

Encyclopedia of
Complexity and Systems Science Series
Editor-in-Chief: Robert A. Meyers

SPRINGER NATURE []
Reference

Tsau-Young Lin · Churn-Jung Liao
Janusz Kacprzyk
Editors

Granular, Fuzzy, and Soft Computing

A Volume in the Encyclopedia of
Complexity and Systems Science,
Second Edition

Encyclopedia of Complexity and Systems Science Series

Editor-in-Chief

Robert A. Meyers, Ramtech Limited, Palm Desert, CA, USA

The Encyclopedia of Complexity and Systems Science series of topical volumes provides an authoritative source for understanding and applying the concepts of complexity theory, together with the tools and measures for analyzing complex systems in all fields of science and engineering. Many phenomena at all scales in science and engineering have the characteristics of complex systems, and can be fully understood only through the transdisciplinary perspectives, theories, and tools of self-organization, synergetics, dynamical systems, turbulence, catastrophes, instabilities, nonlinearity, stochastic processes, chaos, neural networks, cellular automata, adaptive systems, genetic algorithms, and so on. Examples of near-term problems and major unknowns that can be approached through complexity and systems science include: The structure, history and future of the universe; the biological basis of consciousness; the integration of genomics, proteomics and bioinformatics as systems biology; human longevity limits; the limits of computing; sustainability of human societies and life on earth; predictability, dynamics and extent of earthquakes, hurricanes, tsunamis, and other natural disasters; the dynamics of turbulent flows; lasers or fluids in physics, microprocessor design; macromolecular assembly in chemistry and biophysics; brain functions in cognitive neuroscience; climate change; ecosystem management; traffic management; and business cycles. All these seemingly diverse kinds of phenomena and structure formation have a number of important features and underlying structures in common. These deep structural similarities can be exploited to transfer analytical methods and understanding from one field to another. This unique work will extend the influence of complexity and system science to a much wider audience than has been possible to date.

Tsau-Young Lin • Churn-Jung Liao •
Janusz Kacprzyk
Editors

Granular, Fuzzy, and Soft Computing

A Volume in the Encyclopedia of
Complexity and Systems Science,
Second Edition

With 193 Figures and 87 Tables

 Springer

Editors

Tsau-Young Lin
Institute of Data Science
GrC Society
San Jose, CA, USA

Churn-Jung Liao
Inst of Information Science
Academia Sinica
Taipei, Taiwan

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Warsaw, Poland

ISSN 2629-2327 ISSN 2629-2343 (electronic)
Encyclopedia of Complexity and Systems Science Series
ISBN 978-1-0716-2627-6 ISBN 978-1-0716-2628-3 (eBook)
<https://doi.org/10.1007/978-1-0716-2628-3>

© Springer Science+Business Media, LLC, part of Springer Nature 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Series Preface

The Encyclopedia of Complexity and System Science Series is a multivolume authoritative source for understanding and applying the basic tenets of complexity and systems theory as well as the tools and measures for analyzing complex systems in science, engineering, and many areas of social, financial, and business interactions. It is written for an audience of advanced university undergraduate and graduate students, professors, and professionals in a wide range of fields who must manage complexity on scales ranging from the atomic and molecular to the societal and global.

Complex systems are systems that comprise many interacting parts with the ability to generate a new quality of collective behavior through self-organization, e.g., the spontaneous formation of temporal, spatial, or functional structures. They are, therefore, adaptive as they evolve and may contain self-driving feedback loops. Thus, complex systems are much more than a sum of their parts. Complex systems are often characterized as having extreme sensitivity to initial conditions as well as emergent behavior that are not readily predictable or even completely deterministic. The conclusion is that a reductionist (bottom-up) approach is often an incomplete description of a phenomenon. This recognition that the collective behavior of the whole system cannot be simply inferred from the understanding of the behavior of the individual components has led to many new concepts and sophisticated mathematical and modeling tools for application to many scientific, engineering, and societal issues that can be adequately described only in terms of complexity and complex systems.

Examples of grand scientific challenges which can be approached through complexity and systems science include: the structure, history, and future of the universe; the biological basis of consciousness; the true complexity of the genetic makeup and molecular functioning of humans (genetics and epigenetics) and other life forms; human longevity limits; unification of the laws of physics; the dynamics and extent of climate change and the effects of climate change; extending the boundaries of and understanding the theoretical limits of computing; sustainability of life on the earth; workings of the interior of the earth; predictability, dynamics, and extent of earthquakes, tsunamis, and other natural disasters; dynamics of turbulent flows and the motion of granular materials; the structure of atoms as expressed in the Standard Model and the formulation of the Standard Model and gravity into a Unified Theory; the structure of water; control of global infectious diseases; and also evolution and

quantification of (ultimately) human cooperative behavior in politics, economics, business systems, and social interactions. In fact, most of these issues have identified nonlinearities and are beginning to be addressed with nonlinear techniques, e.g., human longevity limits, the Standard Model, climate change, earthquake prediction, workings of the earth's interior, natural disaster prediction, etc.

The individual complex systems mathematical and modeling tools and scientific and engineering applications that comprised the *Encyclopedia of Complexity and Systems Science* are being completely updated, and the majority will be published as individual books edited by experts in each field who are eminent university faculty members. The topics are as follows:

Agent-Based Modeling and Simulation
 Applications of Physics and Mathematics to Social Science
 Cellular Automata, Mathematical Basis of
 Chaos and Complexity in Astrophysics
 Climate Modeling, Global Warming, and Weather Prediction
 Complex Networks and Graph Theory
 Complexity and Nonlinearity in Autonomous Robotics
 Complexity in Computational Chemistry
 Complexity in Earthquakes, Tsunamis, and Volcanoes, and Forecasting and
 Early Warning of Their Hazards
 Computational and Theoretical Nanoscience
 Control and Dynamical Systems
 Data Mining and Knowledge Discovery
 Ecological Complexity
 Ergodic Theory
 Finance and Econometrics
 Fractals and Multifractals
 Game Theory
 Granular Computing
 Intelligent Systems
 Nonlinear Ordinary Differential Equations and Dynamical Systems
 Nonlinear Partial Differential Equations
 Percolation
 Perturbation Theory
 Probability and Statistics in Complex Systems
 Quantum Information Science
 Social Network Analysis
 Soft Computing
 Solitons
 Statistical and Nonlinear Physics
 Synergetics
 System Dynamics
 Systems Biology

Each entry in each of the Series books was selected, and peer reviews were organized by one of our university-based book editors with advice and

consultation provided by our eminent board members and the editor-in-chief. This level of coordination assures that the reader can have a level of confidence in the relevance and accuracy of the information far exceeding than that generally found on the World Wide Web. Accessibility is also a priority, and for this reason, each entry includes a glossary of important terms and a concise definition of the subject. In addition, we are pleased that the mathematical portions of our Encyclopedia have been selected by Math Reviews for indexing in MathSciNet. Also, ACM, the world's largest educational and scientific computing society, recognized our *Computational Complexity: Theory, Techniques, and Applications* book, which contains content taken exclusively from the *Encyclopedia of Complexity and Systems Science*, with an award as one of the notable Computer Science publications. Clearly, we have achieved prominence at a level beyond our expectations, but consistent with the high quality of the content!

Palm Desert, CA, USA
March 2023

Robert A. Meyers
Editor-in-Chief

Volume Preface

The very purpose of this Granular, Fuzzy, and Soft Computing Volume (will be called Encyclopedia) in the Encyclopedia of complexity and system science is to provide the reader with a comprehensive and constructive introduction to some main concepts, research directions, and approaches to deal with imperfect data, information, and knowledge in mathematics, science, and technology. This volume is divided into two parts: the first part will be traditional mathematics based, while the second part will be (new) soft computing based, to which that has been appended with nature-inspired, plus biologically and linguistically motivated, computational paradigms. Let us still name them.

Soft and Nature-Inspired Computing

In general, we are concerned, first, with how to deal with both the traditional uncertainty that is usually represented via probabilities of some events and then dealt with using tools and techniques of probability theory and statistics; this is the classical mathematical approach. Next, which is one of our main topics in this part of the Encyclopedia, we are concerned with imperfect information that is related to imprecision and vagueness of meaning that emerges in virtually all situations in which natural language is employed with its inherent imprecision and vagueness, both from the point of view of semantics and syntax.

Soft computing was coined by Zadeh probably in the early 1980s to describe it as a consortium of the following areas: fuzzy logic (later augmented with rough sets), neural networks, and evolutionary computations. In fact, originally, probabilistic reasoning and machine learning were added, but this had not been followed as these areas concern different problems and use different tools and techniques. It should be noticed that these ingredients of soft computing have a common feature, namely, are meant to model and solve problems for which they do not presuppose to have exact models exemplified by complex models based on differential equations and strict optimization.

Though the term soft computing is still widely used, computational intelligence is getting more and more popular and widely employed, to a large extent due to the existence of the Computational Intelligence Society within the Institute of Electrical and Electronic Engineering (IEEE), the world's

largest engineering organization. There, the mission of computational intelligence is stated as to advance nature-inspired computational paradigms in science and engineering, while the field of interest is said to be the theory, design, application, and development of biologically and linguistically motivated computational paradigms emphasizing neural networks, connectionist systems, genetic algorithms, evolutionary programming, fuzzy systems, and hybrid intelligent systems, to be more specific neural networks, evolutionary computations of all kinds, fuzzy sets and fuzzy logic (with rough sets), ant and swarm type algorithms, “deeper” nature-inspired paradigms like the immunological paradigm, chaotic systems and computations, harmony search algorithms, probabilistic reasoning, etc. These topics expand those initially included in soft computing.

A very important problem in this context is related to some jointly employed tools and techniques. In this respect, we discuss the concept and main properties of the so-called neuro-fuzzy systems, an extremely powerful and successful combination of neural nets and fuzzy sets, and also a combination of imprecision and uncertainty, which is of fuzziness and probability, or – to be more specific – in an important new “statistical” analysis based on fuzzy (imprecise) data. The term soft computing has been appended with new “statistical” analysis. The paper based on such a view is collected in the second part of this volume. Let us move to the details of the first part.

Mathematical Fuzzy Sets and Rough Sets

The theory of fuzzy sets introduced by Zadeh (1965) is here a very popular and powerful apparatus, with both tens and thousands of large world congresses and conferences, prestigious journals, and also many business and industrial applications. Observe that fuzzy sets are defined by very precise objects, called functions, in pure mathematics. So fuzzy sets and fuzzy controls with such an emphasis are included in the first part of this volume. For example, the fuzzy papers presented by Professor Li, who constructed a real hardware equipment for four-stage inverted pendulum (in 2002), and Professor Chang, who coined fuzzy topology, are included in the first part.

A popular alternative to deal with similar aspects of imperfect information, maybe more related to vagueness, missing values, etc., may here also be Pawlak’s (1981) rough set theory, which has also attracted a considerable interest among researchers and scholars resulting also in many papers, books, and respected conference proceedings. At this point, it may be important to observe that the imprecise or vague information in rough set theory can be expressed by precise topological concepts. The upper and lower approximations of a given concept X (a subset in the universe of a rough set theory) are the topological concepts, called closure and interior of X , respectively. The imprecise or vague information collected together is the topological concept of boundary points. So rough set theory leads us naturally into a new horizon, called topological data analysis. Let us name it.

(Generalized) Topological Computing

“Coffee cup and donut are topologically equal?” is a common introduction to the subject of topology. A question for soft computing person is: Can we find the equivalent fuzzy set on the donut that has been defined on the coffee cup?

This topological data analysis started in Poland quietly by Zdzislaw Pawlak. It generates some momentum from two workshops: (1) Rough Set and Soft Computing workshop in 1994 at San Jose and (2) ACM1995 database mining workshop. Note that in his lecture, Professor May (U of Chicago) stated that for a finite set X , the topologies on X are in bijective correspondence with the preorders on X (reflexive and transitive binary relations). So the term topology in topological data analysis includes generalized topology (e.g., Frechet (V) spaces introduced in Sierpinski and Krieger’s book, *General Topology*).

Granular Computing (GRC)

The concept of information granulation has been coined by Zadeh, first in the realm of fuzzy logic but even in Zadeh’s initial view this was much more general, not limited to fuzzy sets (fuzzy logic). Shortly after Zadeh’s pioneering ideas, a new impetus for the development and use of granulation has occurred as a result of the growing popularity of Pawlak’s (1981) rough set theory. In 1988 ACM, Lin motivated by Hao Wang’s approximate proof and goal queries in database community has initiated a research on approximate retrieval in database using generalized topology (called neighborhood system). At that time, we were not aware of the topological nature of rough set theory, but we moved swiftly into it. In 1996 Fall, we visited Berkeley, and the term granular computing to identify a subset of Zadeh’s granular mathematics as his area of research (see T. Y. Lin: ► [“Granular Computing: Practices, Theories, and Future Directions”](#) in this volume).

In 2000, we were invited to organize ICDM2001 and followed it by organizing a few workshops in this series of conferences. During this period, we have established granular computing society and cosponsored IEEE granular computing. At ACM/IEEE WI2008, we were invited to introduce the concept of simplicial complexes of keywords (algebraic topology) into text analysis. In other words, our granular computing is moving into new phases. During 2005–2014, IEEE granular computing conferences have naturally emerged into IEEE bigdata conferences, which were cosponsored jointly by IEEE and GrC -society for a few years. The development of topological data analysis based on rough sets, granular computing, and (generalized) topology began to mature. We have settled (1) some fundamental issues on information flow security (2017), (2) proposed “perfect” fuzzy theory for fuzzy controls (2018), and (3) established the approximate arithmetic for numerical analysis (in this 2022 volume). Granular computing (and generalized topological computing) are flourishing in many areas, such as knowledge engineering, security, social network, information flow analysis, etc.

Soft and Various Forms of Intelligent Computing

The uncertainty of a fuzzy set is derived from how the membership function is introduced, which can, for instance, result from the use of inherently imprecise and vague natural language even though a membership function is a function, which is a perfect object in pure mathematics. Proper tools and techniques are therefore needed and are provided.

First, the main perspective assumed is related to a crucial problem of information granulation that occurs in all kinds of human activities and its related concept of granular computing. Basically, the very concept of an information granule boils down to what has always been done by humans (and then, for instance, other agents mimicking the human behavior) while solving more complex problems.

Generally speaking, information granules are some collections of items, for instance, numbers, which are considered to be, from some point of view, similar or the same due to their similarity, functionality, indistinguishability, coherency, etc. For instance, if we have water temperature that can be from the freezing point (0 °C) to the boiling point (100 °C), then depending on our interest or purpose we can decide that for our problem only some granules are important, exemplified by cold water (e.g., 0–20 °C), lukewarm water (e.g., 26–50 °C), warm water (e.g., 46–70 °C), and hot water (e.g., 71–100 °C). So, the information granulation is via the use of value intervals, which is a commonly employed practice in science and technology, and the granulation process is according to what is needed; for other purposes, the granulation can be different, that is, in this case, the value intervals that are relevant can be different.

A very important aspect of granulation is related to the use of natural language that is the only fully natural means of articulation and communication for the humans beings, and here the use of terms like “cold,” “lukewarm,” “warm,” “hot,” as well as of composite terms like “slightly warmer than lukewarm,” to just mention a few, is an example of granulation. Needless to say that information granulation is commonly used by humans for dealing with and solving all kinds of real-world problems.

In this work, we briefly present a readable and comprehensive account of fuzzy sets and fuzzy logic, both in the sense of basic concepts and properties, and also some tools and techniques to model and process relations between variables that are characterized by the imprecision of values and relations which can be used to describe and model various real-world problems, notably with human beings as key elements.

The concept of information granulation has been coined by Zadeh, first in the realm of fuzzy logic but even in Zadeh’s initial view this was much more general, not limited to fuzzy sets (fuzzy logic). A boost to the broadly perceived granularity of information has for sure been given by another groundbreaking idea of Zadeh termed computing with words that has triggered a huge research interest, notably after Zadeh and Kacprzyk’s (1999) big two-volume editorial project that has collected virtually all developments in the field in that time. These concepts, and then tools and techniques of information granulation and computing with words, have resulted in hundreds

of publications in top journals, conference proceedings, and edited volumes. Moreover, they have been instrumental in many real-world applications and implementations exemplified by fuzzy control, database querying, linguistic summarization, data analytics, etc.

Shortly after Zadeh's pioneering ideas, a new impetus for the development and use of granulation has occurred as a result of a growing popularity of Pawlak's (1981), T.Y. Lin (1996) generalize rough sets in the context of generalized topology (neighborhood systems ACM abstract, 1988) a new research direction of information granulation; see the first part of this volume.

Many important issues related to the broadly perceived information granulation are presented in the Encyclopedia, with a particular emphasis on the aggregation of evidence which occurs in virtually all problems. Computational intelligence is certainly a field that will attract more and more attention.

The research areas mentioned above, that is, fuzzy sets (fuzzy logic), rough sets, and generalized topological computing, will be presented in this Encyclopedia with the purpose of presenting a set of tools and techniques that can be used by interested readers to deal with problems in which information or knowledge that is important is, first of all, imprecise or vague. This can, for instance, result from the use of inherently imprecise and vague natural language. Proper tools and techniques are therefore needed, and brief, yet constructive introductions to fuzzy sets (fuzzy logic), rough sets, and generalized topology are provided.

San Jose, USA
Taipei, Taiwan
Warsaw, Poland
March 2023

Tsau-Young Lin
Churn-Jung Liao
Janusz Kacprzyk
Editors

Contents

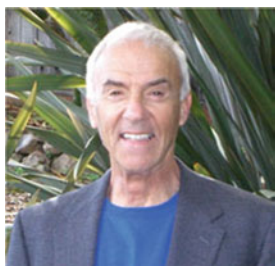
Cooperative Multi-hierarchical Query Answering Systems	1
Zbigniew W. Ras and Agnieszka Dardzinska	
Dependency and Granularity in Data-Mining	7
Shusaku Tsumoto and Shoji Hirano	
Fuzzy Logic	19
Lotfi A. Zadeh	
Fuzzy Probability Theory	51
Michael Beer	
Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions	77
I. Burhan Türkşen	
On Genetic-Fuzzy Data-Mining Techniques	97
Tzung-Pei Hong, Chun-Hao Chen and Vincent S. Tseng	
Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach	117
Salvatore Greco, Benedetto Matarazzo and Roman Słowiński	
Granular Computing, Information Models for	147
Steven A. Demurjian, Alberto De la Rosa Algarin and Rishi Knath Saripalle	
Introduction to Granular Computing	161
Tsau-Young Lin	
Granular Computing and Modeling of the Uncertainty in Quantum Mechanics	167
Kow-Lung Chang	
Philosophical Foundation for Granular Computing	177
Zhengxin Chen	
Granular Computing: Practices, Theories, and Future Directions	199
Tsau-Young Lin	

Principles and Perspectives of Granular Computing	221
Jianchao Han and Nick Cercone	
Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities	239
De Wang, Steve Webb, Kyumin Lee, James Caverlee and Calton Pu	
Granular Model for Data Mining	251
Anita Wasilewska and Ernestina Menasalvas	
Granular Neural Networks	265
Yan-Qing Zhang	
Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems	279
Lech Polkowski	
Multi-Granular Computing and Quotient Structure	311
Ling Zhang and Bo Zhang	
Non-standard Analysis, an Invitation to	323
Wei-Zhe Yang	
Information System Design Using Fuzzy and Rough Set Theory	355
Frederick Petry	
Rough Set Data Analysis	375
Shusaku Tsumoto	
Rule Induction, Missing Attribute Values, and Discretization	389
Jerzy W. Grzymala-Busse	
Social Networks and Granular Computing	401
Churn-Jung Liao	
Rough Sets in Decision-Making	419
Roman Słowiński, Salvatore Greco and Benedetto Matarazzo	
Knowledge Engineering from Email Archives	469
Arvind Srinivasan and Chien-Chung Chan	
Fuzzy Similarity for Parallel Function Computation Model	487
Chin-Liang Chang	
Mereology	501
Hsing-chien Tsai	
Variable Precision Approximations of Rough Sets	517
Wojciech Ziarko	
Algebraic Models and Granular Computing	529
Anita Wasilewska	

Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm	551
Sadaaki Miyamoto	
Fuzzy System Representations of Stochastic Systems	563
Hong-Xing Li	
Function Approximation Property of Fuzzy Systems and Its Error Analysis	593
Hong-Xing Li	
Issues in Access Control and Privacy for Big Data	615
Sylvia L. Osborn, Daniel Servos and Motahera Shermin	
Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment	625
Thomas H. Hinke and Derek Shaw	
Security with Privacy	639
Elisa Bertino	
Privacy Preservation in Big Data Analytics	649
Yu-Chuan Tsai, Shyue-Liang Wang and Tzung-Pei Hong	
Access Control for XML Big Data Applications	671
Alberto De la Rosa Algarin, Steven A. Demurjian and Eric Jackson	
A Theory of Approximate Arithmetic for Numerical Analysis: A Mathematical Foundation	693
Tsau-Young Lin	
Interval Computing	701
Ralph Baker Kearfott	
Aggregation Operators and Soft Computing	711
Vicenç Torra	
Evolving Fuzzy Systems	725
Plamen Angelov	
Fuzzy Logic, Type-2 and Uncertainty	743
Robert I. John and Jerry M. Mendel	
Fuzzy Optimization	755
Weldon A. Lodwick and Elizabeth A. Untiedt	
Foundations of Fuzzy Sets Theory	793
Janusz Kacprzyk	
Hybrid Soft Computing Models for Systems Modeling and Control	821
Oscar Castillo and Patricia Melin	
Neuro-fuzzy Systems	843
Leszek Rutkowski, Krzysztof Cpalka, Robert Nowicki, Agata Pokropinska and Rafal Scherer	

Possibility Theory	859
Didier Dubois and Henri Prade	
Rough Sets: Foundations and Perspectives	877
James F. Peters, Andrzej Skowron and Jaroslaw Stepaniuk	
Introduction to Soft Computing	891
Janusz Kacprzyk	
Statistics with Imprecise Data	895
María Ángeles Gil and Olgierd Hryniewicz	
Some Remarks on the Foundations of Numerical Analysis	911
Ralph Baker Kearfott	
Index	917

About the Editor-in-Chief



Robert A. Meyers

President: RAMTECH Limited

Manager, Chemical Process Technology, TRW Inc.

Postdoctoral Fellow: California Institute of Technology

Ph.D. Chemistry, University of California at Los Angeles

B.A. Chemistry, California State University, San Diego

Dr. Meyers was manager of Energy and Environmental Projects at TRW (now Northrop Grumman) in Redondo Beach, CA, and is now president of RAMTECH Limited. He is coinventor of the gravimelt process for desulfurization and demineralization of coal for air pollution and water pollution control and was manager of the Department of Energy project leading to the construction and successful operation of a first-of-a-kind Gravimelt Process Integrated Test Plant. He is the inventor of and was project manager for the DOE-sponsored Magnetohydrodynamics Seed Regeneration Project, which has resulted in the construction and successful operation of a pilot plant for the production of potassium formate, a chemical utilized for plasma electricity generation and air pollution control. He also managed TRW efforts in magnetohydrodynamics electricity generating combustor and plasma channel development. He managed the pilot-scale DoE project for determining the hydrodynamics of synthetic fuels. He is a coinventor of several thermooxidative stable polymers which have achieved commercial success as the GE PEI, Upjohn Polyimides, and Rhone-Poulenc bismaleimide resins. He has also managed projects

for photochemistry, chemical lasers, flue gas scrubbing, oil shale analysis and refining, petroleum analysis and refining, global change measurement from space satellites, analysis and mitigation (carbon dioxide and ozone), hydrometallurgical refining, soil and hazardous waste remediation, novel polymers synthesis, modeling of the economics of space transportation systems, space rigidizable structures, and chemiluminescence-based devices.

He is a senior member of the American Institute of Chemical Engineers, member of the American Physical Society, and member of the American Chemical Society and has served on the UCLA Chemistry Department Advisory Board. He was a member of the joint USA-Russia working group on air pollution control and the EPA-sponsored Waste Reduction Institute for Scientists and Engineers.

Dr. Meyers has more than 20 patents and 50 technical papers in the fields of photochemistry, pollution control, inorganic reactions, organic reactions, luminescence phenomena, and polymers. He has published in primary literature journals including *Science* and the *Journal of the American Chemical Society* and is listed in *Who's Who in America* and *Who's Who in the World*. His scientific achievements have been reviewed in feature articles in the popular press in publications such as *The New York Times Science Supplement* and *The Wall Street Journal* as well as more specialized publications such as *Chemical Engineering* and *Coal Age*. A public service film was produced by the Environmental Protection Agency on Dr. Meyers' chemical desulfurization invention for air pollution control.

Dr. Meyers is the author or editor-in-chief of a wide range of technical books including the *Handbook of Chemical Production Processes*; the *Handbook of Synfuels Technology*; the *Handbook of Petroleum Refining Processes*, now in the fourth edition; the *Handbook of Petrochemical Production Processes* (McGraw-Hill), now in a second edition; the *Handbook of Energy Technology and Economics*, published by John Wiley & Sons; *Coal Structure*, published by Academic Press; and *Coal Desulfurization* as well as the *Coal Handbook* published by Marcel Dekker. He served as chairman of the advisory board for *A Guide to Nuclear Power Technology*, published by John Wiley & Sons,

which won the Association of American Publishers Award as the best book in technology and engineering. He also served as editor-in-chief of three editions of the *Elsevier Encyclopedia of Physical Science and Technology*. Most recently, he serves as editor-in-chief of the *Encyclopedia of Analytical Chemistry* as well as *Reviews in Cell Biology and Molecular Medicine* and a book series of the same name both published by John Wiley & Sons. In addition, he currently serves as editor-in-chief of two Springer Nature book series: *Encyclopedia of Complexity and Systems Science* and *Encyclopedia of Sustainability Science and Technology*.

About the Volume Editors



Tsau-Young Lin holds a PhD in mathematics from Yale University. Prof. Lin started his career as a mathematician, but his work has been extended to computer science. Let us quote his latest work in mathematics that involves computer science. In his book entitled *Numerical Analysis*, L. Ridgway Scott stated “there is not a well-established mathematical model for finite-precision arithmetic” (Ch1, p1, third bullet). Prof. Lin’s latest work is to build such a model. The solution is included in this Encyclopedia.

Prof. Lin has published and edited well over 250 papers and 10 books. In addition, at Silicon Valley, he has done a considerable amount of local services. For example, Prof. Lin has invited many “Nobel Laureates,” such as John McCarthy, Steve Smale, and Lotfi Zadeh, to speak at IEEE local chapter and started a series of local IEEE conferences called “New Frontier in Computing.”

In 1996, Fall Prof. Lin took his sabbatical leave at BISC (Berkeley Initiated Soft Computing). To name his research project. Prof. Lin coined the term granular computing from Zadeh’s granular mathematics. There were two families of granules in Lin’s mind. The first one is called infinitesimal in ancient time, which is implicitly mentioned in Zadeh’s paper (1979). This concept is called topology, and we will include its extension called the system of neighborhoods of a point in generalized topology (e.g., Frechet (V) spaces in Ch 1. Sec 1, General Topology by Sierpinski and Krieger). The second one is a simplex, which technically is a primitive in algebraic topology and represents a primitive concept in a document set. In his invited talk at WI2008, Prof. Lin presented his first vision, since then much progress has been achieved

(see IEEE BigData 2016). The community of topological data analysis seems maturing; we foresee much progress in near future.

Prof. Lin has been a frontrunner in various fields in data analysis and mining. He is the Founding President of the Rough Set (topological data analysis) Society and President of the GrC Society (generalized point set and algebraical topological data analysis). He organized the very database mining workshop in CSC95 (earlier than KDD), the first ICDM conference, and the first US Web Intelligence conference. Prof. Lin also started and cosponsored the first few IEEE bigdata conferences and the full series of IEEE granular computing. Prof. Lin's areas of expertise include data science (deep data analysis, data/knowledge mining, uncertainty), generalized-topological data analysis (algebraic, geometric, point set topology), fuzzy control oriented fuzzy set theory (IEEE SMC 2018), secure information flow (IEEE BigData 2017), and very fast frequent item set mining (IEEE BigData 2016). Though granular computing is basically derived from Zadeh's granular mathematics, Prof. Lin has been taking mathematical routes.

Let us turn back the clock. In the early 1980s, Prof. Lin decided to move west. He gave up his tenure and joined the defense industry. To gain a feel for a real large-scale parallel computing system, Prof. Lin volunteered to build a "compiler" for a language that is designed for such a system. Such industry experiences have played important roles in his later research. He settled the Peterson conjecture in Petri net theory, which is a mathematical model of parallel computing. Based on his first-hand observations on information flows within defense contractors, including the actual actions of trusted subjects, Prof. Lin published a sequence of security papers in IFIP11.3 security conferences, including Chinese wall security policy paper in ACSAC 1989. During the time of this research, Prof. Lin first joined the California State University at Northridge and then at San Jose (had been part time at Los Angeles). Prof. Lin retired after 50 years of academic life, but he started his second career immediately. As we have mentioned above, the starting point of this second career is a shining one.



Churn-Jung Liao received a BS, an MS, and a PhD in computer science and information engineering from National Taiwan University in 1985, 1987, and 1992, respectively. He then joined the Institute of Information Science, Academia Sinica, Taiwan, and is currently a tenured full research fellow. His major research interest includes applied logic, artificial intelligence, and reasoning about uncertainty.



Janusz Kacprzyk is Professor of Computer Science at the Systems Research Institute, Polish Academy of Sciences, WIT – Warsaw School of Information Technology, and Chongqing Three Gorges University, Wanzhou, Chongqing, China, and Professor of Automatic Control at PIAP – Industrial Institute of Automation and Measurements in Warsaw, Poland. He is Honorary Foreign Professor at the Department of Mathematics, Yulin Normal University, Xinjiang, China. He is Full Member of the Polish Academy of Sciences; Member of Academia Europaea; European Academy of Sciences and Arts; European Academy of Sciences; Foreign Member of the: Bulgarian Academy of Sciences, Spanish Royal Academy of Economic and Financial Sciences (RACEF), Finnish Society of Sciences and Letters, Flemish Royal Academy of Belgium of Sciences and the Arts (KVAB), National Academy of Sciences of Ukraine, and Lithuanian Academy of Sciences. He was awarded seven honorary doctorates. He is Fellow of IEEE, IET, IFSA, EurAI, IFIP, AAIA, I2CICC, and SMIA.

His main research interests include the use of modern computation computational and artificial intelligence tools, notably fuzzy logic, in systems science, decision making, optimization, control, data analysis, and data mining, with applications in mobile robotics, systems modeling, ICT, etc.

He authored 7 books, (co)edited more than 150 volumes, and (co)authored more than 650 papers, including ca. 150 in journals indexed by the WoS. He was listed in 2020 and 2021 “World’s 2% Top Scientists” by Stanford

University, Elsevier (Scopus), and SciTech Strategies and published in *PLOS Biology* journal.

He is the editor-in-chief of 8 book series at Springer, and of 2 journals, and is on the editorial boards of ca. 40 journals. He is President of the Polish Operational and Systems Research Society and Past President of International Fuzzy Systems Association.

Contributors

Plamen Angelov Intelligent Systems Research Laboratory, Digital Signal Processing Research Group, Communication Systems Department, Lancaster University, Lancaster, UK

Michael Beer Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany

Elisa Bertino CS Department, Purdue University, West Lafayette, IN, USA

Oscar Castillo Division of Graduate Studies and Research, Tijuana Institute of Technology, Tijuana, Mexico

James Caverlee Department of Computer Science, Texas A&M University, College Station, TX, USA

Nick Cercone Faculty of Science and Engineering, York University, Toronto, ON, Canada

Chien-Chung Chan Department of Computer Science, The University of Akron, Akron, OH, USA

Chin-Liang Chang Nicesoft Corporation, Austin, TX, USA

Kow-Lung Chang Physics Department, National Taiwan University, Taipei, Taiwan

Chun-Hao Chen Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan

Zhengxin Chen Department of Computer Science, University of Nebraska at Omaha, Omaha, NE, USA

Krzysztof Cpalka Department of Computer Engineering, Czestochowa University of Technology, Czestochowa, Poland

Department of Artificial Intelligence, Academy of Humanities and Economics, Lodz, Poland

Agnieszka Dardzinska Department of Computer Science, Białystok Technical University, Białystok, Poland

Alberto De la Rosa Algarin Cynnovative, LLC, Arlington, VA, USA

Steven A. Demurjian Department of Computer Science and Engineering,
University of Connecticut, Storrs, CT, USA

Didier Dubois IRT-CNRS, Universite Paul Sabatier, Toulouse, France

María Ángeles Gil Faculty of Sciences, University of Oviedo, Oviedo, Spain

Salvatore Greco Department of Economics and Business, University of
Catania, Catania, Italy

Jerzy W. Grzymala-Busse Department of Electrical Engineering and Com-
puter Science, University of Kansas, Lawrence, KS, USA

Department of Expert Systems and Artificial Intelligence, University of Infor-
mation Technology and Management, Rzeszow, Poland

Jianchao Han Department of Computer Science, California State University
Dominguez Hills, Carson, CA, USA

Thomas H. Hinke Moffett Field, CA, USA

Shoji Hirano Department of Medical Informatics, Shimane University,
School of Medicine, Enya-cho Izumo City, Shimane, Japan

Tzung-Pei Hong Department of Computer Science and Information Engi-
neering, National University of Kaohsiung, Kaohsiung, Taiwan

Olgierd Hryniewicz Systems Research Institute, Warsaw, Poland

Eric Jackson Department of Civil and Environmental Engineering, Univer-
sity of Connecticut, Storrs, CT, USA

Robert I. John Centre for Computational Intelligence, School of Computing,
De Montfort University, Leicester, UK

Janusz Kacprzyk Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Ralph Baker Kearfott Department of Mathematics, University of Louisiana,
Lafayette, LA, USA

Kyumin Lee Department of Computer Science, Utah State University,
Logan, UT, USA

Hong-Xing Li School of Control Science and Engineering, Dalian Univer-
sity of Technology, Dalian, P. R. China

Churn-Jung Liao Institute of Information Science, Academia Sinica, Taipei,
Taiwan

Tsau-Young Lin Institute of Data Science, GrC Society, San Jose, CA, USA

Weldon A. Lodwick Department of Mathematical Sciences, University of
Colorado Denver, Denver, CO, USA

Benedetto Matarazzo Department of Economics and Business, University
of Catania, Catania, Italy

Patricia Melin Division of Graduate Studies and Research, Tijuana Institute of Technology, Tijuana, Mexico

Ernestina Menasalvas Departamento de Lenguajes y Sistemas Informaticos, Facultad de Informatica, Madrid, Spain

Jerry M. Mendel Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA

Sadaaki Miyamoto University of Tsukuba, Tsukuba, Japan

Robert Nowicki Departament of Computer Engineering, Czestochowa University of Technology, Czestochowa, Poland

Sylvia L. Osborn Department of Computer Science, Western University, London, ON, Canada

James F. Peters Computational Intelligence Laboratory, University of Manitoba, Winnipeg, MB, Canada

Frederick Petry Naval Research Lab, Stennis Space Center, Mississippi, MS, USA

Agata Pokropinska Institute of Mathematics and Computer Science, Jan Dlugosz University, Czestochowa, Poland

Lech Polkowski Mathematics and Computer Science, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland

Henri Prade IRT-CNRS, Universite Paul Sabatier, Toulouse, France

Calton Pu College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Zbigniew W. Ras Department of Computer Science, University of North Carolina, Charlotte, NC, USA

Leszek Rutkowski Departament of Computer Engineering, Czestochowa University of Technology, Czestochowa, Poland

Rishi Knath Saripalle University of Connecticut, Storrs, CT, USA

Rafal Scherer Departament of Computer Engineering, Czestochowa University of Technology, Czestochowa, Poland

Daniel Servos Department of Computer Science, Western University, London, ON, Canada

Derek Shaw NASA Advanced Supercomputing Division, NASA Ames Research Center, Moffett Field, CA, USA

Motahera Shermin Department of Computer Science, Western University, London, ON, Canada

Andrzej Skowron Institute of Mathematics, Warsaw University, Warsaw, Poland

Roman Słowiński Institute of Computing Science, Poznań University of Technology, Poznań, Poland

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Arvind Srinivasan SplitByte Inc., Los Gatos, CA, USA

Jarosław Stepaniuk Computer Science, Białystok University of Technology, Białystok, Poland

Vicenç Torra Institut d'Investigació en Intel·ligència Artificial – CSIC, Bellaterra, Spain

Hsing-chien Tsai Department of Philosophy, National Chung-Cheng University, Chia-Yi, Taiwan

Yu-Chuan Tsai Library and Information Center, National University of Kaohsiung, Kaohsiung, Taiwan

Vincent S. Tseng Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Shusaku Tsumoto Department of Medical Informatics, Faculty of Medicine, Shimane University, Enya-cho Izumo City, Shimane, Japan

I. Burhan Türkşen Head Department of Industrial Engineering, TOBB-ETÜ, (Economics and Technology University of the Union of Turkish Chambers and Commodity Exchanges), Ankara, Turkey

Elizabeth A. Untiedt Department of Mathematical Sciences, University of Colorado Denver, Denver, CO, USA

De Wang College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Shyue-Liang Wang Department of Information Management, National University of Kaohsiung, Kaohsiung, Taiwan

Anita Wasilewska Computer Science Department, Stony Brook University, Stony Brook, NY, USA

Steve Webb College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Wei-Zhe Yang Department of Mathematics, National Taiwan University, Taipei, Taiwan

Lotfi A. Zadeh Department of EECS, University of California, Berkeley, CA, USA

Bo Zhang Department of Computer Science, State Key Lab of Intelligent Technology and Systems, Tsinghua University, Beijing, China

Ling Zhang Artificial Intelligence Institute, Anhui University, Hefei, Anhui, China

Yan-Qing Zhang Department of Computer Science, Georgia State University, Atlanta, GA, USA

Wojciech Ziarko Department of Computer Science, University of Regina, Regina, SK, Canada

Cooperative Multi-hierarchical Query Answering Systems

Zbigniew W. Ras¹ and Agnieszka Dardzinska²

¹Department of Computer Science, University of North Carolina, Charlotte, NC, USA

²Department of Computer Science, Białystok Technical University, Białystok, Poland

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Multi-hierarchical Decision System
 Cooperative Query Answering
 Future Directions
 Bibliography

Glossary

Autonomous information system An autonomous information system is an information system existing as an independent entity.

Intelligent query answering Intelligent query answering is an enhancement of query-answering into a sort of intelligent system (capable of being adapted or molded). Such systems should be able to interpret incorrectly posed questions and compose an answer not necessarily reflecting precisely what is directly referred to by the question, but rather reflecting what the intermediary understands to be the intention linked with the question.

Knowledge base Knowledge base is a collection of rules defined as expressions written in predicate calculus. These rules have a form of associations between conjuncts of values of attributes.

Ontology Ontology is an explicit formal specification of how to represent objects, concepts

and other entities that are assumed to exist in some area of interest and relationships holding among them. Systems that share the same ontology are able to communicate about the domain of discourse without necessarily operating on a globally shared theory. A system commits to ontology if its observable actions are consistent with the definitions in the ontology.

Semantics The meaning of expressions written in some language as opposed to their syntax which describes how symbols may be combined independently of their meaning.

Definition of the Subject

One way to make a Query Answering System (QAS) intelligent is to assume the hierarchical structure of their attributes. Such systems have been investigated by Cuppens and Demolombe (1988), Gal and Minker (1988), and Gaasterland et al. (1992), and they are called cooperative. Queries submitted to them are built, in a classical way, from values of attributes describing objects in an information system S and from two-argument functors “and,” “or.” Instead of “or,” we use the symbol “+”. Instead of “and,” we use the symbol “*”. Let us assume that QAS is associated with an information system S . Now, if query q submitted to QAS fails, then any attribute value listed in q can be generalized and the number of objects supporting q in S may increase. In cooperative systems, these generalizations are controlled either by users (Gal and Minker 1988), or by methods based on knowledge discovery (Muslea 2004). Conceptually, a similar approach has been proposed by Lin (1989). He defines a neighborhood of an attribute value which we can interpret as its generalization (or its parent in the corresponding hierarchical attribute structure). When query fails, then the query answering system is trying to replace values in a query by new values from their corresponding neighborhoods. QAS for S can also collaborate

and exchange knowledge with other information systems. In all such cases, it is called intelligent. In papers (Ras and Dardzinska 2004, 2006) the query answering strategy was based on a guided process of knowledge (rules) extraction and knowledge exchange among systems. Knowledge extracted from information systems collaborating with S was used to construct new attributes in S and/or impute null or hidden values of attributes in S . This way we do not only enlarge the set of queries which QAS can successfully answer but also increase the overall number of retrieved objects and their confidence. Some attributes in S can be distinguished. We usually call them decision attributes. Their values represent concepts which can be defined in terms of the remaining attributes in S , called classification attributes. Query languages for such information systems are built only from values of decision attributes and from two-argument functors “+”, “*” (Ras et al. 2007). The semantics of queries is defined in terms of semantics of values of classification attributes. Precision and recall of QAS is strictly dependent on the support and confidence of the classifiers used to define queries.

Introduction

Responses by QAS to submitted queries do not always contain the information desired and although they may be logically correct, can sometimes be misleading. Research in the area of intelligent query answering rectifies these problems. The classical approach is based on a cooperative method called relaxation for expanding an information system and related to it queries (Cuppens and Demolombe 1988; Gal and Minker 1988). The relaxation method expands the scope of a query by relaxing the constraints implicit in the query. This allows QAS to return answers related to the original query as well as the literal answers which may be of interest to the user.

This paper concentrates on multi-hierarchical decision systems which are defined as information systems with several hierarchical distinguished attributes called decision attributes. Their values are used to build queries. We give the theoretical

framework for modeling such systems and its corresponding query languages. Standard interpretation and the classifier-based interpretation of queries are introduced and used to model the quality (precision, recall) of QAS.

Multi-hierarchical Decision System

In this section we introduce the notion of a multi-hierarchical decision system S and the query language built from atomic expressions containing only values of the decision attributes in S . Classifier-based semantics and the standard semantics of queries in S are proposed. The set of objects X in S is defined as the interpretation domain for both semantics. Standard semantics identifies all objects in X which should be retrieved by a query. Classifier-based semantics gives weighted sets of objects which are retrieved by queries. The notion of precision and recall of QAS in the proposed setting is introduced. We use only rule-based classifiers to define the classifier-based semantics. By improving the confidence and support of the classifiers we improve the precision and recall of QAS.

Definition 1 By a multi-hierarchical decision system we mean a triple $S = (X, A \cup D, V)$, where X is a nonempty, finite set of objects, $D = \{d[i]: 1 \leq i \leq k\}$ is a set of hierarchical decision attributes, A is a nonempty finite set of classification attributes, and $V = \bigcup \{V_a: a \in A \cup D\}$ is a set of their values.

We assume that: V_a, V_b are disjoint for any $a, b \in A \cup D$, such that $a \neq b$, $a: X \rightarrow V_a$ is a partial function for every $a \in A \cup D$.

Definition 2 By a set of decision queries (d -queries) for S we mean a least set T_D such that:

- $0, 1 \in T_D$,
- if $w \in \bigcup \{V_a: a \in D\}$, then $w, \sim w \in T_D$,
- if $t_1, t_2 \in T_D$, then $(t_1 + t_2), (t_1 * t_2) \in T_D$.

Definition 3 Decision query t is called simple if $t = t_1 * t_2 * \dots * t_n$ and

$$(\forall j \in \{1, 2, \dots, n\})[(t_j \in \cup \{V_a : a \in D\}) \\ \vee (t_j = \sim w \wedge w \in \cup \{V_a : a \in D\})].$$

Definition 4 By a set of classification terms (c-terms) for S we mean a least set T_C such that:

- $0, 1 \in T_C$,
- if $w \in \cup \{V_a : a \in A\}$, then $w, \sim w \in T_C$,
- $t_1, t_2 \in T_C$, then $(t_1 + t_2), (t_1 * t_2) \in T_C$.

Definition 5 Classification term t is called simple if $t = t_1 * t_2 * \dots * t_n$ and

$$(\forall j \in \{1, 2, \dots, n\})[(t_j \in \cup \{V_a : a \in A\}) \\ \vee (t_j = \sim w \wedge w \in \cup \{V_a : a \in A\})].$$

Definition 6 By a classification rule we mean any expression of the form $[t_1 \rightarrow t_2]$, where t_1 is a simple classification term and t_2 is a simple decision query.

Definition 7 Semantics M_S of c-terms in $S = (X, A \cup D, V)$ is defined in a standard way as follows:

- $M_S(0) = 0, M_S(1) = X$,
- $M_S(w) = \{x \in X : w = a(x)\}$ for any $w \in V_a, a \in A$,

$$M_S(\sim w) = \{x \in X : (\exists v \in V_a)[v = a(x) \& \\ v \neq w]\} \text{ for any } w \in V_a, a \in A,$$

- if t_1, t_2 are terms, then

$$M_S(t_1 + t_2) = M_S(t_1) \cup M_S(t_2), \\ M_S(t_1 * t_2) = M_S(t_1) \cap M_S(t_2).$$

Now, we introduce the notation used for values of decision attributes. Assume that the term $d[i]$ also denotes the first granularity level of a hierarchical decision attribute $d[i]$. The set $\{d[i, 1], d[i, 2], d[i, 3], \dots\}$ represents the values of attribute $d[i]$ at its second granularity level. The set $\{d[i, 1, 1], d[i, 1, 2], \dots, d[i, 1, n_i]\}$ represents the values of attribute d at its third granularity level, right below the node $d[i, 1]$. We assume here that the value $d[i, 1]$ can be refined to any value from $\{d[i, 1, 1], d[i, 1, 2], \dots, d[i, 1, n_i]\}$, if necessary.

Similarly, the set $\{d[i, 3, 1, 3, 1], d[i, 3, 1, 3, 2], d[i, 3, 1, 3, 3], d[i, 3, 1, 3, 4]\}$ represents the values of attribute d at its fourth granularity level which are finer than the value $d[i, 3, 1, 3]$.

Now, let us assume that a rule-based classifier is used to extract rules describing simple decision queries in S . We denote that classifier by **RC**. The definition of semantics N_S of c-terms is **RC** independent whereas the definition of semantics M_S of d-queries is **RC** dependent.

Definition 8 Classifier-based semantics M_S of d-queries in $S = (X, A \cup D, V)$ is defined as follows: If t is a simple d-query in S and $\{r_j = [t_j \rightarrow t] : j \in J_t\}$ is a set of all rules defining t which are extracted from S by classifier **RC**, then $M_S(t) = \{(x, p_x) : (\exists j \in J_t)(x \in M_S(t_j)[p_x = \Sigma \{\text{conf}(j) \cdot \text{sup}(j) : x \in M_S(t_j) \& j \in J_t\} / \Sigma \{\text{sup}(j) : x \in M_S(t_j) \& j \in J_t\}])\}$, where $\text{conf}(j)$, $\text{sup}(j)$ denote the confidence and the support of the rule $[t_j \rightarrow t]$, correspondingly.

Definition 9 Attribute $d[j_1, j_2, \dots, j_n]$ in $S = (X, A \cup D, V)$, is dependent on $d[i_1, i_2, \dots, i_k]$ in S , if one of the following conditions hold:

$$n \leq k \& (\forall m \leq n)[i_m = j_m],$$

$$n > k \& (\forall m \leq k)[i_m = j_m].$$

Otherwise, $d[j_1, j_2, \dots, j_n]$ is called independent from $d[i_1, i_2, \dots, i_k]$ in S .

Example 1 The attribute value $d[2, 3, 1, 2]$ is dependent on the attribute value $d[2, 3, 1, 2, 5, 3]$. Also, $d[2, 3, 1, 2, 5, 3, 2, 4]$ is dependent on $d[2, 3, 1, 2, 5, 3]$.

Definition 10 Let $S = (X, A \cup \{d[1], d[2], \dots, d[k]\}, V)$, $w \in V_{d[i]}$, and $IV_{d[i]}$ be the set of all attribute values in $V_{d[i]}$ which are independent from w .

Standard semantics N_S of d-queries in S is defined as follows:

- $N_S(0) = 0, N_S(1) = X$,
- if $w \in V_{d[i]}$, then $N_S(w) = \{x \in X : d[i](x) = w\}$, for any $1 \leq i \leq k$

- if $w \in V_{d[i]}$, then $N_S(\sim w) = \{x \in X : (\exists v \in IV_{d[i]}[d[i](x) = v]) \text{ for any } 1 \leq i \leq k\}$
- if t_1, t_2 are terms, then

$$N_S(t_1 + t_2) = N_S(t_1) \cup N_S(t_2),$$

$$N_S(t_1 * t_2) = N_S(t_1) \cap N_S(t_2).$$

Definition 11 Let $S = (X, A \cup D, V)$, t is a d -query in S , $N_S(t)$ is its meaning under standard semantics, and $M_S(t)$ is its meaning under classifier-based semantics. Assume that $N_S(t) = X_1 \cup Y_1$, where $X_1 = \{x_i, i \in I_1\}$, $Y_1 = \{y_i, i \in I_2\}$, Assume also that $M_S(t) = \{(x_i, p_i) : i \in I_1\} \cup \{(z_i, q_i) : i \in I_3\}$ and $\{y_i, i \in I_2\} \cap \{z_i, i \in I_3\} = \emptyset$.

By precision of a classifier-based semantics M_S on a d -query t , we mean

$$\text{Prec}(M_S, t) = [\Sigma\{p_i : i \in I_1\} + \Sigma\{(1 - q_i) : i \in I_3\}] / [\text{card}(I_1) + \text{card}(I_3)].$$

By recall of a classifier-based semantics M_S on a d -query t , we mean

$$\text{Rec}(M_S, t) = [\Sigma\{p_i : i \in I_1\}] / [\text{card}(I_1) + \text{card}(I_2)].$$

Example 2 Assume that $N_S(t) = \{x_1, x_2, x_3, x_4\}$, $M_S(t) = \{(x_1, p_1), (x_2, p_2), (x_5, p_5), (x_6, p_6)\}$ Then:

$$\text{Prec}(M_S, t) = [p_1 + p_2 + (1 - p_5) + (1 - p_6)] / 4,$$

$$\text{Rec}(M_S, t) = [p_1 + p_2] / 4.$$

$$\text{Prec}(M_S, t) = [p_1 + p_2 + (1 - p_5) + (1 - p_6)] / 4,$$

$$\text{Rec}(M_S, t) = [p_1 + p_2] / 4.$$

Cooperative Query Answering

There are cases when classical Query Answering Systems fail to return any answer to a d -query q but still a satisfactory answer can be found. For instance, let us assume that in a multi-hierarchical decision system $S = (X, A \cup D, V)$, where $D = \{d[1], d[2], \dots, d[k]\}$, there is no

single object whose description matches the query q . Assuming that a distance measure between objects in S is defined, then by generalizing q , we may identify objects in S whose descriptions are closest to the description of q . This problem is similar to the problem when the granularity of an attribute value used in a query q is finer than the granularity of the corresponding attribute used in S . By replacing such attribute values in q by more general values used in S , we may retrieve objects from S which satisfy q .

Definition 12 The distance δ_S between two attribute values $d[j_1, j_2, \dots, j_n]$, $d[i_1, i_2, \dots, i_m]$ in $S = (X, A \cup D, V)$, where $j_1 = i_1$, $p \geq 1$, is defined as follows:

1. if $[j_1, j_2, \dots, j_p] = [i_1, i_2, \dots, i_p]$ and $j_{p+1} \neq i_{p+1}$, then $\delta_S[d[j_1, j_2, \dots, j_n], d[i_1, i_2, \dots, i_m]] = 1/2^{p-1}$.
2. if $n \leq m$ and $[j_1, j_2, \dots, j_n] = [i_1, i_2, \dots, i_n]$, then $\delta_S[d[j_1, j_2, \dots, j_n], d[i_1, i_2, \dots, i_m]] = 1/2^n$.

The second condition, in the above definition, represents the average case between the best and the worst case.

Example 3 Following the above definition of the distance measure, we get:

$$\delta_S[d[2, 3, 2, 4], d[2, 3, 2, 5, 1]] = 1/4$$

$$\delta_S[d[2, 3, 2, 4], d[2, 3, 2]] = 1/8$$

Let us assume that $q = q(a[3, 1, 3, 2], b[1], c[2])$ is a d -query submitted to S . The notation $q(a[3, 1, 3, 2], b[1], c[2])$ means that q is built from $a[3, 1, 3, 2]$, $b[1]$, $c[2]$ which are the atomic attribute values in S . Additionally, we assume that attribute a is not only hierarchical but also it is ordered. It basically means that the difference between the values $a[3, 1, 3, 2]$ and $a[3, 1, 3, 3]$ is smaller than between the values $a[3, 1, 3, 2]$ and $a[3, 1, 3, 4]$. Also, the difference between any two elements in $\{a[3, 1, 3, 1], a[3, 1, 3, 2], a[3, 1, 3, 3], a[3, 1, 3, 4]\}$ is smaller than between $a[3, 1, 3]$ and $a[3, 1, 2]$.

Now, we outline a possible strategy which QAS can follow to solve q . Clearly, the best solution for answering q is to identify objects in S which precisely match the d -query submitted by the user. If it fails, we try to identify objects which match d -query $q(a[3, 1, 3], b[1], c[2])$. If we succeed, then we try d -queries $q(a[3, 1, 3, 1], b[1], c[2])$ and $q(a[3, 1, 3, 3], b[1], c[2])$. If we fail, then we should succeed with $q(a[3, 1, 3, 4], b[1], c[2])$. If we fail with $q(a[3, 1, 3], b[1], c[2])$, then we try $q(a[3, 1], b[1], c[2])$ and so on.

To present this cooperative strategy in a more precise way, we use an example and start with a very simple dataset. Namely, we assume that S has four decision attributes which belong to the set $\{a, b, c, d\}$. System S contains only four objects listed in Table 1.

Now, we assume that d -query $q = a[1, 2] * b[2] * c[1, 1] * d[3, 1, 1]$ is submitted to the multi-hierarchical decision system S (see Table 1). Clearly, q fails in S .

Jointly with q , also a threshold value for a minimum support can be supplied as a part of a d -query. This threshold gives the minimal number of objects that need to be returned as an answer to q . When QAS fails to answer q , the nearest objects satisfying q have to be identified.

The algorithm for finding these objects is based on the following steps:

If QAS fails to identify a sufficient number of objects satisfying q in S , then the generalization process starts. We can generalize either attribute a or d . Since the value $d[3, 1, 2]$ has a lower granularity level than $a[1, 1]$ then we generalize $d[3, 1, 2]$ getting a new query $q_1 = a[1, 2] * b[2] * c[1, 1] * d[3, 1]$. But q_1 still fails in S . Now, we generalize $a[1, 1]$ getting a new query $q_2 = a[1] * b[2] * c[1, 1] * d[3, 1]$. Objects x_1, x_2 are the only objects in S which support q_2 .

If the user is only interested in one object satisfying the query q , then we need to identify which object in $\{x_1, x_2\}$ has a distance closer to q .

Clearly,

$$\begin{aligned}\delta_S[q, x_1] &= \delta_S[[a[1, 2], b[2], c[1, 1], d[3, 1, 1]], [a[1], \\ &\quad b[2], c[1, 1], d[3]]] \\ &= 1/4 + 0 + 0 + 1/4 = 1/2, \\ \delta_S[q, x_2] &= \delta_S[[a[1, 2], b[2], c[1, 1], d[3, 1, 1]], \\ &\quad [a[1, 1], b[2, 1], c[1, 1, 1], d[3, 1, 2]]] \\ &= 1/4 + 1/4 + 1/8 + 1/8 = 3/4,\end{aligned}$$

which means x_1 is the winning object.

Note that the cooperative strategy only identifies objects satisfying d -queries and it identifies objects to be returned by QAS to the user. The confidence assigned to these objects depends on the classifier **RC**.

Future Directions

We have introduced the notion of system-based semantics and user-based semantics of queries. User-based semantics are associated with the indexing of objects by a user which is time consuming and unrealistic for very large sets of data. System-based semantics are associated with automatic indexing of objects in X which strictly depends on the support and confidence of classifiers and depends on the precision and recall of a query answering system. The quality of classifiers can be improved by a proper enlargement of the set X and the set of features describing them which differentiate the real-life objects from the same semantic domain as X in a better way. An example, for instance, is given in (Ras et al. 2007). The quality of a query answering system can be

Cooperative Multi-hierarchical Query Answering Systems, Table 1 Multi-hierarchical decision system S

X	e	f	g	...	...	a	b	c	d
x_1	$e[1]$	$f[1]$	...	...	...	$a[1]$	$b[2]$	$c[1, 1]$	$d[3]$
x_2	$e[2]$	$f[1]$	...	...	...	$a[1, 1]$	$b[2, 1]$	$c[1, 1, 1]$	$d[3, 1, 2]$
x_3	$e[2]$	$f[1]$	...	...	...	$a[1, 1, 1]$	$b[2, 2, 1]$	$c[2, 2]$	$d[1]$
x_4	$e[1]$	$f[2]$	...	...	...	$a[2]$	$b[2, 2]$	$c[1, 1]$	$d[1, 1]$

improved by its cooperativeness. Both precision and recall of QAS is increased if no-answer queries are replaced by generalized queries which are answered by QAS on a higher granularity level than the initial level of queries submitted by users. Assuming that the system is distributed, the quality of QAS for multi-hierarchical decision system S can also be improved through collaboration among sites (Ras and Dardzinska 2004, 2006).

The key concept of intelligent QAS based on collaboration among sites is to generate global knowledge through knowledge sharing. Each site develops knowledge independently which is used jointly to produce global knowledge. Assume that two sites S_1 and S_2 accept the same ontology of their attributes and share their knowledge in order to solve a user query successfully. Also, assume that one of the attributes at site S_1 is confidential. The confidential data in S_1 can be hidden by replacing them with null values. However, users at S_1 may treat them as missing data and reconstruct them with the knowledge extracted from S_2 (Im and Ras 2007). The vulnerability illustrated in this example shows that a security-aware data management is an essential component for any intelligent QAS to ensure data confidentiality.

Bibliography

- Chmielewski MR, Grzymala-Busse JW, Peterson NW (1993) The rule induction system LERS – a version for personal computers. *Found Comput. Decis Sci* 18(3-4):181–212
- Chu W, Yang H, Chiang K, Minock M, Chow G, Larson C (1996) Cobase: a scalable and extensible cooperative information system. *J Intell Inf Syst* 6(2/3):223–259
- Cuppens F, Demolombe R (1988) Cooperative answering: a methodology to provide intelligent access to databases. In: *Proceeding of the second international conference on expert database systems*, pp 333–353
- Gaasterland T (1997) Cooperative answering through controlled query relaxation. *IEEE Expert* 12(5):48–59
- Gaasterland T, Godfrey P, Minker J (1992) Relaxation as a platform for cooperative answering. *J Intell Inf Syst* 1(3):293–321
- Gal A, Minker J (1988) Informative and cooperative answers in databases using integrity constraints. *Natural language understanding and logic programming*. North Holland, pp 277–300
- Giannotti F, Manco G (2002) Integrating data mining with intelligent query answering. In: *Logics in artificial intelligence, Lecture notes in computer science*, vol 2424. Springer, Berlin, pp 517–520
- Godfrey P (1993) Minimization in cooperative response to failing database queries. *Int J Coop Inf Syst* 6(2):95–149
- Guarino N (1998) *Formal ontology in information systems*. IOS Press, Amsterdam
- Im S, Ras ZW (2007) Protection of sensitive data based on reducts in a distributed knowledge discovery system. In: *Proceedings of the international conference on multimedia and ubiquitous engineering (MUE 2007)*, Seoul. IEEE Computer Society, pp 762–766
- Lin TY (1989) Neighborhood systems and approximation in relational databases and knowledge bases. *Proceedings of the fourth international symposium on methodologies for intelligent systems. Poster session program*, Oak Ridge National Laboratory, ORNL/DSRD-24, pp 75–86
- Muslea I (2004) Machine learning for online query relaxation. *Proceedings of KDD-2004*, Seattle. ACM, pp 246–255
- Pawlak Z (1981) Information systems – theoretical foundations. *Inf Syst J* 6:205–218
- Ras ZW, Dardzinska A (2004) Ontology based distributed autonomous knowledge systems. *Inf Syst Int J* 29(1):47–58
- Ras ZW, Dardzinska A (2006) Solving failing queries through cooperation and collaboration, special issue on web resources access. *World Wide Web J* 9(2):173–186
- Ras ZW, Zhang X, Lewis R (2007) MIRAI: Multi-hierarchical, FS-tree based music information retrieval system. In: Kryszkiewicz M et al (eds) *Proceedings of RSEISP, LNAI*, vol 4585. Springer, Berlin, pp 80–89

Dependency and Granularity in Data-Mining

Shusaku Tsumoto¹ and Shoji Hirano²

¹Department of Medical Informatics, Faculty of Medicine, Shimane University, Enya-cho Izumo City, Shimane, Japan

²Department of Medical Informatics, Shimane University, School of Medicine, Enya-cho Izumo City, Shimane, Japan

Article Outline

Definition of the Subject
Introduction
Contingency Table from Rough Sets
Rank of Contingency Table (2×2)
Rank of Contingency Table ($m \times n$)
Rank and Degree of Dependence
Conclusion
Bibliography

Definition of the Subject

The degree of granularity of a contingency table is closely related with that of dependence of contingency tables. We investigate these relations from the viewpoints of determinantal divisors and determinants. From the results of determinantal divisors, it seems that the divisors provide information on the degree of dependencies between the matrix of all elements and its submatrices and that an increase in degree of granularity may lead to an increase in dependency. However, another approach shows that a constraint on the sample size of a contingency table is very strong, which leads to an evaluation formula in which an increase of degree of granularity gives a decrease of dependency.

Introduction

Independence (dependence) is a very important concept in data mining, especially for feature selection. In rough sets (Pawlak 1991), if two attribute-value pairs, say $[c = 0]$ and $[d = 0]$ are independent, their supporting sets, denoted by C and D do not have an overlapping region ($C \cap D = \phi$), which means that an attribute independent to a given target concept may not appear in the classification rule for the concept.

This idea is also frequently used in other rule discovery methods: let us consider deterministic rules, described as *if-then* rules, which can be viewed as classic propositions ($C \rightarrow D$). From the set-theoretical point of view, a set of examples supporting the conditional part of a deterministic rule, denoted by C , is a subset of a set whose examples belong to the consequence part, denoted by D . That is, the relation $C \subseteq D$ holds and deterministic rules are supported only by positive examples in a dataset (Tsumoto 2000).

When such a subset relation is not satisfied, indeterministic rules can be defined as if-then rules with probabilistic information (Tsumoto and Tanaka 1996). From the set-theoretical point of view, C is not a subset, but closely overlapped with D . That is, the relations $C \cap D \neq \phi$ and $|C \cap D| / |C| \geq \delta$ will hold in this case. (The threshold δ is the degree of the closeness of overlapping sets, which will be given by domain experts. For more information, please refer to section “Rank of Contingency Table (2×2).”) Thus, probabilistic rules are supported by a large number of positive examples and a small number of negative examples.

On the other hand, in a probabilistic context, independence of two attributes means that one attribute (a_1) will not influence the occurrence of the other attribute (a_2), which is formulated as $p(a_2|a_1) = p(a_2)$.

Although independence is a very important concept, it has not been fully and formally investigated as a relation between two attributes. Tsumoto introduces linear algebra into formal

analysis of a contingency table (Tsumoto 2003). The results give the following interesting results: First, a contingency table can be viewed as comparison between two attributes with respect to information granularity. Second, algebra is a key point of analysis of this table. A contingency table can be viewed as a matrix and several operations and ideas of matrix theory are introduced into the analysis of the contingency table. Especially for the degree of independence, rank plays a very important role in extracting a probabilistic model from a given contingency table.

This paper presents a further investigation into the degree of independence of a contingency matrix.

Intuitively and empirically, when two attributes have many values, the dependence between these two attributes becomes low. However, from the results of determinantal divisors, it seems that the divisors provide information on the degree of dependencies between the matrix of all elements and its submatrices and that an increase in degree of granularity may lead to an increase in dependency. The key of the resolution of these conflicts is to consider constraints on the sample size.

In this paper we show that a constraint on the sample size of a contingency table is very strong, which leads to the evaluation formula in which the increase of degree of granularity gives a decrease of dependency. The paper is organized as follows: Section “Contingency Table from Rough Sets” shows preliminaries. Section “Rank of Contingency Table (2×2)” discusses the former results. Section “Rank of Contingency Table ($m \times n$)” gives the relations between rank and submatrices of a matrix. Finally, section “Degree of Granularity and Dependence” concludes this paper.

Contingency Table from Rough Sets

Notations

In the subsequent sections, the following notation, introduced in (Skowron and Grzymala-Busse 1994), is adopted: Let U denote a nonempty, finite set called the universe and A denote a nonempty, finite set of attributes, that is, $a : U \rightarrow V_a$ for $a \in A$, where V_a is called the domain of a . Then, a decision table is defined as an information

system, $A = (U, A \cup \{\mathcal{D}\})$, where $\{\mathcal{D}\}$ is a set of given decision attributes. The atomic formulas over $B \subseteq A \cup \{\mathcal{D}\}$ and V are expressions of the form $[a = v]$, called descriptors over B , where $a \in B$ and $v \in V_a$. The set $F(B, V)$ of formulas over B is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction and negation. For each $f \in F(B, V)$, f_A denotes the meaning of f in A , that is, the set of all objects in U with property f , defined inductively as follows:

1. If f is of the form $[a = v]$ then, $f_A = \{s \in U \mid a(s) = v\}$.
2. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \cup g_A$; $(\neg f)_A = U - f_A$.

Contingency Table ($m \times n$)

Definition 1 Let R_1 and R_2 denote multinomial attributes in an attribute space A which have m and n values.

A contingency table is a table of a set described by the following formulas: $|[R_1 = A_j]_A|$, $|[R_2 = B_i]_A|$, $|[R_1 = A_j \wedge R_2 = B_i]_A|$, $|[R_1 = A_1 \wedge R_1 = A_2 \wedge \dots \wedge R_1 = A_m]_A|$, $|[R_2 = B_1 \wedge R_2 = A_2 \wedge \dots \wedge R_2 = A_n]_A|$ and $|U|$ ($i = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, m$).

This table is arranged into the form shown in Table 1, where: $|[R_1 = A_j]_A| = \sum_{i=1}^m x_{ji} = x_{.j}$, $|[R_2 = B_i]_A| = \sum_{j=1}^n x_{ji} = x_{i.}$, $|[R_1 = A_j \wedge R_2 = B_i]_A| = x_{ij}$, $|U| = N = x_{..}$ ($i = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, m$).

Example 1 Let us consider an information table shown in Table 2.

The relationship between b and e can be examined by using the corresponding contingency table. First, the frequencies of four elementary relations,

Dependency and Granularity in Data-Mining, Table 1 Contingency table ($m \times n$)

	A_1	A_2	$\dots$	A_n	Sum
B_1	X_{11}	X_{12}	$\dots$	X_{1n}	$X_{1.}$
B_2	X_{21}	X_{22}	$\dots$	X_{2n}	$X_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
B_m	X_{m1}	X_{m2}	$\dots$	X_{mn}	$X_{m.}$
Sum	$X_{.1}$	$X_{.2}$	$\dots$	$X_{.n}$	$X_{..} = U = N$

Dependency and Granularity in Data-Mining,
Table 2 A small dataset

a	b	c	d	e
1	0	0	0	1
0	0	1	1	1
0	1	2	2	0
1	1	1	2	1
0	0	2	1	0

Dependency and Granularity in Data-Mining,
Table 3 Corresponding contingency table

	$b = 0$	$b = 1$	
$e = 0$	1	1	2
$e = 1$	2	1	3
	3	2	5

called *marginal distributions*, are counted: $[b = 0]$, $[b = 1]$, $[e = 0]$, and $[e = 1]$. Then, the frequencies of four kinds of conjunction are counted: $[b = 0] \wedge [e = 0]$, $[b = 0] \wedge [e = 1]$, $[b = 1] \wedge [e = 0]$, and $[b = 1] \wedge [e = 1]$.

Then, the following contingency table is obtained (Table 3).

From this table, accuracy and coverage for $[b = 0] \rightarrow [e = 0]$ are obtained as $1/(1 + 2) = 1/3$ and $1/(1 + 1) = 1/2$.

One of the important observations from granular computing is that a contingency table shows the relations between two attributes with respect to intersection of their supporting sets. For example, in Table 3, both b and e have two different partitions of the universe and the table gives the relation between b and e with respect to the intersection of supporting sets. It is easy to see that this idea can be extended into an $n \times n$ contingency table, which can be viewed as an $n \times n$ -matrix. When two attributes have a different number of equivalence classes, the situation may be a little complicated. But, in this case, due to knowledge about linear algebra, we have only to consider the attribute which has a smaller number of equivalence classes and the surplus number of equivalence classes of the attributes with larger number of equivalence classes can be projected into other partitions. In other words, an $m \times n$ matrix or contingency table includes a projection from one attribute to the other one.

Rank of Contingency Table (2×2)

Preliminaries

Definition 2 A corresponding matrix $C_{T_{a,b}}$ is defined as a matrix the elements of which are equal to the value of the corresponding contingency table $T_{a,b}$ of two attributes a and b , except for marginal values.

Definition 3 The rank of a table is defined as the rank of its corresponding matrix. The maximum value of the rank is equal to the size of (square) matrix, denoted by r .

Example 2 Let the table given in Table 3 be defined as $T_{b,e}$. Then, $C_{T_{b,e}}$ is:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}.$$

Since the determinant of $C_{T_{b,e}}$ $\det(C_{T_{b,e}})$ is not equal to 0, the rank of $C_{T_{b,e}}$ is equal to 2. It is the maximum value ($r = 2$), so b and e are statistically dependent.

Independence When the Table Is 2×2

From the application of linear algebra, several results are obtained. (The proof is omitted). First, it is assumed that a contingency table is given as $m = 2$, $n = 2$ in Table 1. Then the corresponding matrix ($C_{T_{R_1,R_2}}$) is given as:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}.$$

Proposition 1 The determinant of $\det(C_{T_{R_1,R_2}})$ is equal to $|x_{11}x_{22} - x_{12}x_{21}|$.

Proposition 2 The rank will be:

$$\text{rank} = \begin{cases} 2, & \text{if } \det(C_{T_{R_1,R_2}}) \neq 0 \\ 1, & \text{if } \det(C_{T_{R_1,R_2}}) = 0. \end{cases}$$

If the rank of $\det(C_{T_{b,e}})$ is equal to 1, according to the theorems of linear algebra, it is obtained that

one row or column will be represented by the other column. That is,

Proposition 3 Let r_1 and r_2 denote the rows of the corresponding matrix of a given 2×2 table, $C_{T_{b,c}}$. That is,

$$r_1 = (x_{11}, x_{12}), \quad r_2 = (x_{21}, x_{22}).$$

Then, r_1 can be represented by r_2 : $r_1 = kr_2$, where k is given as:

$$k = \frac{x_{11}}{x_{21}} = \frac{x_{12}}{x_{22}} = \frac{x_1}{x_2}.$$

From this proposition, the following theorem is obtained.

Theorem 1 If the rank of the corresponding matrix is 1, then two attributes in a given contingency table are statistically independent.

Thus,

$$\text{rank} = \begin{cases} 2, & \text{dependent} \\ 1, & \text{statistical independent.} \end{cases}$$

Rank of Contingency Table ($m \times n$)

In the case of a general square matrix, the results in the 2×2 contingency table can be extended. It is especially important to observe that conventional statistical independence is supported only when the rank of the corresponding matrix is equal to 1. Let us consider the contingency table of c and a in Table 2, which is obtained as follows. Thus, the corresponding matrix of this table is:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

whose determinant is equal to 0. It is clear that its rank is 2.

It is interesting to see that if the case of $[d = 0]$ is removed, then the rank of the corresponding matrix is equal to 1 and two rows are equal. Thus, if the value space of d into $\{1, 2\}$ is restricted,

Dependency and Granularity in Data-Mining, Table 4 Contingency table for a and c

	$a = 0$	$a = 1$	
$c = 0$	0	1	1
$c = 1$	1	1	2
$c = 2$	2	0	2
	3	2	5

then c and d are statistically independent. This relation is called *contextual independence* (Butz 2002), which is related to conditional independence.

However, another type of weak independence is observed: let us consider the contingency table of a and c . The table is obtained as Table 4.

Its corresponding matrix is:

$$C_{T_{a,c}} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 0 \end{pmatrix},$$

Since the corresponding matrix is not square, the determinant is not defined. But it is easy to see that the rank of this matrix is two. In this case, even removing any attribute-value pair from the table will not generate statistical independence. Finally, the relation between rank and independence in a $n \times n$ contingency table is obtained.

Theorem 2 Let the corresponding matrix of a given contingency table be a square $n \times n$ matrix. If the rank of the corresponding matrix is 1, then two attributes in a given contingency table are statistically independent. If the rank of the corresponding matrix is n , then two attributes in a given contingency table are dependent. Otherwise, two attributes are contextually dependent, which means that several conditional probabilities can be represented by a linear combination of conditional probabilities. Thus,

$$\text{rank} = \begin{cases} n & \text{dependent} \\ 2, \dots, n-1 & \text{contextual independent} \\ 1 & \text{statistical independent.} \end{cases}$$

This theorem can be generalized into $m \times n$ matrix. If the corresponding matrix of a given contingency table is not square and of the form $m \times n$, then its rank is at most $\min(m, n)$.

For example, since $C_{T_{a,c}}$ is 3×2 , the rank is at most 2. Actually, from the calculation of subdeterminants shown in the next section, this matrix has a rank of 2.

Theorem 3 *Let the corresponding matrix of a given contingency table be an $m \times n$ matrix. The*

rank of this matrix is less than $\min(m, n)$. If the rank of the corresponding matrix is 1, then two attributes in a given contingency table are statistically independent. If the rank of the corresponding matrix is n , then two attributes in a given contingency table are dependent. Otherwise, two attributes are contextually dependent, which means that several conditional probabilities can be represented by a linear combination of conditional probabilities. Thus,

$$\text{rank} = \begin{cases} \min(m, n) & \text{dependent} \\ 2, \dots, \min(m, n) - 1 & \text{contextual independent} \\ 1 & \text{statistical independent.} \end{cases}$$

In the cases of $m \neq n$, we need a discussion on submatrix and subdeterminant in the next section.

Rank and Degree of Dependence

Submatrix and Subdeterminant

The next interest is the structure of a corresponding matrix with $1 \leq \text{rank} \leq n - 1$. First, let us define a submatrix (a subtable) and subdeterminant.

Definition 4 Let A denote a corresponding matrix of a given contingency table ($m \times n$). A corresponding submatrix $A_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r}$ is defined as a matrix which is given by an intersection of r rows and s columns of A ($i_1 < i_2 < \dots < i_r$, $j_1 < j_2 < \dots < j_s$).

Definition 5 A subdeterminant of A is defined as a determinant of a submatrix $A_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r}$, which is denoted by $\det(A_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r})$.

Let us consider the contingency table given as Table 1. Then, a subtable for $A_{j_1 j_2 \dots j_s}^{i_1 i_2 \dots i_r}$ is given as Table 5.

Rank and Subdeterminant

Let δ_{ij} denote a co-factor of a_{ij} in a square corresponding matrix of A . Then,

$$\Delta_{ij} = (-1)^{i+j} \det(A_{1,2,\dots,i-1,i+1,\dots,n}^{1,2,\dots,j-1,j+1,\dots,n}).$$

It is notable that a co-factor is a special type of submatrix, where only the i th-row and j -column are removed from a original matrix.

By the use of co-factors, the determinant of A is defined as:

$$\det(A) = \sum_{j=1}^n a_{ij} \Delta_{ij},$$

which is called *Laplace expansion*.

Dependency and Granularity in Data-Mining, Table 5 A subtable ($r \times s$)

	A_{j_1}	A_{j_2}	$\dots$	A_{j_r}	Sum
B_{i_1}	$X_{i_1 j_1}$	$X_{i_1 j_2}$	$\dots$	$X_{i_1 j_r}$	$X_{i_1 \cdot}$
B_{i_2}	$X_{i_2 j_1}$	$X_{i_2 j_2}$	$\dots$	$X_{i_2 j_r}$	$X_{i_2 \cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
B_{i_r}	$X_{i_r j_1}$	$X_{i_r j_2}$	$\dots$	$X_{i_r j_r}$	$X_{i_r \cdot}$
Sum	$X_{\cdot 1}$	$X_{\cdot 2}$	$\dots$	$X_{\cdot n}$	$X_{\cdot \cdot} = U = N$

From this representation, if $\det(A)$ is not equal to 0, then $\Delta_{ij} \neq 0$ for $\{a_{i1}, a_{i2}, \dots, a_{in}\}$ which are not equal to 0. Thus, the following proposition is obtained.

Proposition 4 $\det(A)$ is not equal to 0 if at least one co-factor of $a_{ij}(\neq 0)$, Δ_{ij} is not equal to 0.

It is notable that the above definition of a determinant gives the relation between an original matrix A and its submatrices (co-factors). Since cofactors give a square matrix of size $n - 1$, the above proposition gives the relation between a matrix of size n and submatrices of size $n - 1$. In the same way, we can discuss the relation between a corresponding matrix of size n and submatrices of size $r(1 \leq r < n - 1)$.

Rank and Submatrix

Let us assume that corresponding matrix and submatrix are square ($n \times n$ and $r \times r$, respectively).

Theorem 4 If the rank of a corresponding matrix of size $n \times n$ is equal to r , the determinant of at least one submatrix of size $r \times r$ is not equal to 0. That is, there exists a submatrix $A_{j_1 j_2 \dots j_r}^{i_1 i_2 \dots i_r}$, which satisfies $\det(A_{j_1 j_2 \dots j_r}^{i_1 i_2 \dots i_r}) \neq 0$.

Corollary 1 If the rank of a corresponding matrix of size $n \times n$ is equal to r , all the determinants of the submatrices whose number of columns and rows are larger than $r + 1(\leq n)$ are equal to 0.

Example 3 Let us consider the corresponding matrix mentioned in the above section, $C_{T_{ac}}$.

The submatrices of this matrix are:

$$\begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Since all the subdeterminants are not equal to 0, the rank of this corresponding matrix is equal to 2.

Example 4 Let us consider the following corresponding matrix:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

The determinant of A is:

$$\begin{aligned} \det(A) &= 1 \times (-1)^{1+1} \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} \\ &\quad + 2 \times (-1)^{1+2} \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} \\ &\quad + 3 \times (-1)^{1+3} \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1 \times (-3) + 2 \times 6 + 3 \times (-3) = 0. \end{aligned}$$

Thus, the rank of A is smaller than 2. All the subdeterminants of A are:

$$\begin{aligned} \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} &= -3, & \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} &= -6, \\ \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} &= -3, & \det \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix} &= -6, \\ \det \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix} &= -12, & \det \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix} &= -6, \\ \det \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} &= -3, & \det \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} &= -6, \\ \det \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} &= -3. \end{aligned}$$

Since all the subdeterminants of A are not equal to 0, the rank of A is equal to 2.

Actually, since.

$$(4 \ 5 \ 6) = \frac{1}{2} \{ (1 \ 2 \ 3) + (7 \ 8 \ 9) \},$$

and $(7 \ 8 \ 9)$ cannot be represented by $k(1 \ 2 \ 3)$ (k : integer), the rank of this matrix is equal to 2.

Thus, one attribute-value pair is statistically dependent on other two pairs, statistically independent of the other attribute. In other words, if

two pairs are fixed, the one remaining attribute-value pair will be statistically independently determined.

Determinantal Divisors

From the subdeterminants of all submatrices of size 2, all the subdeterminants of a corresponding matrix has the greatest common divisor, equal to 3.

From the recursive definition of the determinants, it is show that the subdeterminants of size $r + 1$ will have the greatest common divisor of the subdeterminants of size r as a divisor. Thus,

Theorem 5 Let $d_k(A)$ denote the greatest common divisor of all the subdeterminants of size k , $\det(A_{j_1 i_1 \dots j_r i_r}^{i_1 i_2 \dots i_k})$. $d_1(A), d_2(A), \dots, d_n(A)$ are called *determinantal divisors*. From the definition of Laplace expansion,

$$d_k(A) | d_{k+1}(A).$$

In the example of the above subsection, $d_1(A) = 1, d_2(A) = 3$ and $d_3(A) = 0$.

Example 5 Let us consider $C_{T_{a,c}}$ as an example. $d_1(C_{T_{a,c}}) = 1$ and $d_2(C_{T_{a,c}}) = 1$.

Example 6 Let us consider the following corresponding matrix:

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 11 & 9 \end{pmatrix}.$$

Calculation gives: $d_1(B) = 1, d_2(B) = 3$ and $d_3(B) = 18$.

It is notable that a simple change of a corresponding matrix gives a significant change to the determinant, which suggests a change of structure with regard to dependence/independence.

The relation between $d_k(A)$ gives an interesting constraint.

Proposition 5 Since $d_k(A) | d_{k+1}(A)$, the sequence of the divisors is a monotonically increasing one:

$$d_1(A) \leq d_2(A) \leq \dots \leq d_r(A),$$

where r denotes the rank of A .

The sequence of B illustrates this: $1 < 3 < 18$.

Let us define a ratio of $d_k(A)$ to $d_{k-1}(A)$, called *elementary divisors*, where C denotes a corresponding matrix and $k \leq \text{rank } A$:

$$e_k(C) = \frac{d_k(C)}{d_{k-1}(C)} (d_0(C) = 0).$$

Elementary divisors may give the increase of dependency between two attributes. For example, $e_1(B) = 1, e_2(B) = 3$, and $e_3(B) = 6$. Thus, a transition from 2×2 to 3×3 have a higher impact on the dependency of two attributes.

It is trivial to see that $\det(B) = e_1 e_2 e_3$, which can be viewed as a decomposition of the determinant of a corresponding matrix.

Divisors and Degree of Dependence

Since the determinant can be viewed as the degree of dependence, this result is very important. If values of all the subdeterminants (size r) are very small (nearly equal to 0) and $d_r(A) \simeq 1$, then the values of the subdeterminants (size $r + 1$) are very small. This property may hold until the r reaches the rank of the corresponding matrix. Thus, the sequence of the divisors of a corresponding matrix gives a hidden structure of a contingency table.

Also, these results show that $d_1(A)$ and $d_2(A)$ are very important to estimate the rank of a corresponding matrix. Since $d_1(A)$ is given only by the greatest common divisor of all the elements of A , $d_2(A)$ are much more important components. This also intuitively suggests that the subdeterminants of A with size 2 are principal components of a corresponding matrix from the viewpoint of statistical dependence.

Recall that statistical independence of two attributes is equivalent to a corresponding matrix with a rank of 1. A matrix of rank 2 gives a

context-dependent independence, which means three values of two attributes are independent, but two values of two attributes are dependent.

Subdeterminants and Degree of Dependence

Since the determinants give the degree of dependence, the degree of dependence can be evaluated by the values of subdeterminants.

For the above examples (A), since

$$\det \begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix} = -12$$

gives the maximum value, the first and the third attribute-value pairs for two attributes are dependent on each other.

On the other hand, concerning B , since

$$\det \begin{pmatrix} 2 & 3 \\ 11 & 9 \end{pmatrix} = -15$$

gives the maximum value, the second and the third attribute-value pairs for two attributes are dependent on each other.

This discussion can be extended into the dependency between attribute-value pairs and a corresponding attribute. Let us consider 3×2 submatrices of A , which removes one column of the matrix

$$A_1 = \begin{pmatrix} 2 & 3 \\ 5 & 6 \\ 8 & 9 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 3 \\ 4 & 6 \\ 7 & 9 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix}.$$

From the discussions in the above subsections, a set of the subdeterminants of 2×2 submatrices of A_j , denoted by D_{A_j} are obtained as:

$$\begin{aligned} D_{A_1} &= \{-3, -6, -3\} \\ D_{A_2} &= \{-12, -6, -6\} \\ D_{A_3} &= \{-3, -6, -3\}. \end{aligned}$$

Thus, the first and third attribute value pairs are more dependent than the second value pairs, concerning the classification of attributes for the rows.

Elementary Divisors and Elementary Transformation

Let us define the following three elementary (row/column) transformations of a corresponding matrix:

1. Exchange two rows (columns), i_0 and j_0 ($P(i_0, j_0)$).
2. Multiply -1 to a row (column) i_0 ($T(i_0; -1)$).
3. Multiply t to a row (column) $j_0(i_0)$ and add it to a row $i_0(j_0)$. ($W(i_0, j_0, t)$).

Then, three transformations have several interesting characteristics.

Proposition 6 *Matrices corresponding to three elementary transformations are regular.*

Proposition 7 *Three elementary transformations do not change the rank of a corresponding matrix.*

Proposition 8 *Let $\tilde{A}$ denote a matrix transformed by finite steps of three operations. Then,*

$$\text{rank } \tilde{A} = \text{rank } A, \quad d_r(\tilde{A}) = d_r(A),$$

where r denotes the rank of matrix A .

Then, from the application of linear algebra, the following interesting result is obtained.

Theorem 6 *With the finite steps of elementary transformations, a given corresponding matrix is transformed into*

$$\tilde{A} = \left(\begin{array}{ccc|c} e_1 & & & 0 \\ & e_2 & & \\ & & \ddots & \\ & & & e_r \\ & 0 & & 0 \end{array} \right)$$

where $e_j = \frac{d_j(A)}{d_{j-1}(A)}$ ($d_0(A) = 1$) and r denotes the rank of a corresponding matrix. Then, the determinant is decomposed into the product of e_j .

$$d_r(\tilde{A}) = d_r(A) = e_1 e_2 \dots e_r.$$

Degree of Granularity and Dependence

From Theorem 6, it seems that the increase of the degree of granularity gives that of the dependence between two attributes.

However, our empirical observations are different from the above intuitive analysis. Thus, there should be a strong constraint which suppresses the above effects on the degree of granularity.

Let us assume that the determinant of a given contingency matrix gives the degree of the dependence of the matrix. Then, from the application of linear algebra, we obtain the following theorem.

Theorem 7 *Let A denote an $n \times n$ contingency matrix, which includes N samples. If the rank of A is equal to n , then there exists a matrix B ($n \times n$) which satisfies.*

$$BA = \begin{pmatrix} \rho_1 & & & \\ & \rho_2 & & \\ & & \ddots & \\ & & & \rho_n \end{pmatrix} = P,$$

where $\rho_1 + \rho_2 + \dots + \rho_n = N$.

It is notable that the value of determinants of P is larger than A :

$$\det A \leq \det P.$$

Example 7 Let us consider B as an example (Example 6). Let C denote the orthogonal matrix for transformation of B . Since the cardinality of B is equal to 48, the diagonal matrix which gives the maximum determinant is equal to:

$$\begin{pmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{pmatrix}.$$

On the other hand, the determinant of B is equal to 18. Thus, $\det B = 18 < 16^3 = 4096$. Then, C is obtained from the following equation:

$$C \times \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 11 & 9 \end{pmatrix} = \begin{pmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{pmatrix}.$$

Thus,

$$C = \begin{pmatrix} -56/3 & 40/3 & -8/3 \\ 16/3 & -32/3 & 16/3 \\ 8 & 8/3 & -8/3 \end{pmatrix}.$$

It is notable that the determinant of C is equal to 2048/9. Also, since $\det B = 18$, we do not have any diagonal matrix whose determinant is equal to 18 and whose sum of all the elements is equal to 48.

It is easy to see that the transformed matrix P has a very nice property for calculating the determinant.

Proposition 9 *The determinant of the transformed matrix P is equal to the multiplication of ρ_1 to ρ_n . That is,*

$$\det P = \rho_1 \rho_2 \dots \rho_n.$$

Then, the following constraint will be have the special meaning:

$$\rho_1 + \rho_2 + \dots + \rho_n = N, \quad (1)$$

because the following inequality holds in general:

$$\frac{\rho_1 + \rho_2 + \dots + \rho_n}{n} \geq \sqrt[n]{\rho_1 \rho_2 \dots \rho_n} \quad (2)$$

where the equality holds when $\rho_1 = \rho_2 = \dots = \rho_n$.

Since the above inequality can be transformed into:

$$\rho_1 \rho_2 \dots \rho_n \leq \left(\frac{\rho_1 + \rho_2 + \dots + \rho_n}{n} \right)^n,$$

the following inequality is obtained:

$$\begin{aligned} \det P &= \rho_1 \rho_2 \dots \rho_n \\ &\leq \left(\frac{\rho_1 + \rho_2 + \dots + \rho_n}{n} \right)^n, \end{aligned} \quad (3)$$

where the equality holds when $\rho_1 = \rho_2 = \dots = \rho_n$.

From the Theorem 7 and Eq. (1), the following theorem is obtained.

Theorem 8 When a contingency matrix A holds $AB = P$, where P is a diagonal matrix, the following inequality holds:

$$\det A \leq \left(\frac{N}{n}\right)^n.$$

Proof

$$\begin{aligned} \det A &= \det(PB^{-1}) \\ &\leq \det P \\ &= \rho_1 \rho_2 \cdots \rho_n \\ &\leq \left(\frac{\rho_1 + \rho_2 + \cdots + \rho_n}{n}\right)^n = \left(\frac{N}{n}\right)^n, \end{aligned} \quad (4)$$

where the former equality holds when $\det B^{-1} = \det B = 1$ and the latter equality holds when $\rho_1 = \rho_2 = \cdots = \rho_n = \frac{N}{n}$. $\square$

Example 8 Let us consider the following contingency matrices D and E :

$$D = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 11 & 9 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$E = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 10 & 9 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The numbers of examples of D and E are 49 and 48, respectively, which can be comparable to that of B . Then, from Theorem 8,

$$\det D = 18 < (49/4)^4 = \frac{5764801}{256} \sim 22518$$

$$\det E = 12 < (48/4)^4 = 20736.$$

Thus, the maximum value of the determinant of A is at most $\left(\frac{N}{n}\right)^n$. Since N is constant for the given matrix A , the degree of dependence will decrease very rapidly when n becomes very large. That is,

$$\det A \sim n^{-n}.$$

Thus,

Corollary 2 The determinant of A will converge to 0 when n increases to infinity.

$$\lim_{n \rightarrow \infty} \det A = 0.$$

This result suggests that when the degree of granularity becomes higher, the degree of dependence will become lower due to constraints on the sample size.

However, it is notable that N/n is very important. If N is very large, a rapid decrease will be observed when N is close to n . Even when N is 48 as shown in Example 8, $n = 3, 4$ may give a strong dependency between two attributes. For the behavior of $(N/n)^n$, we can apply the technique of real analysis, which will be our future work.

Conclusion

In this paper, a contingency table is interpreted from the viewpoint of granular computing and statistical independence. Matrix algebra is a key point of the analysis of a contingency table and the degree of independence, and rank plays a very important role in extracting a probabilistic model. From the correspondence between contingency table and matrix, the following results are obtained: First, the value of determinants gives the degree of dependency between attribute-value pairs for a set of submatrices with the same size. Second, from the characteristics of the determinants, the larger the rank a corresponding matrix has, the more the two attributes are dependent. This result is shown by a monotonicity of a sequence of determinantal divisors. Third, elementary divisors give a decomposition of the determinant of a corresponding matrix. Finally, the constraint on the sample size of a contingency table is very strong, which leads to an evaluation formula in which an increase of degree of granularity gives the decrease of dependency.

Acknowledgments This work was supported by the Grant-in-Aid for Scientific Research (13131208) on Priority Areas (No.759) “Implementation of Active Mining in the Era of Information Flood” by the Ministry of Education, Science, Culture, Sports, Science and Technology of Japan.

Bibliography

- Butz C (2002) Exploiting contextual independencies in web search and user profiling. In: Proceedings of World Congress on Computational Intelligence (WCCI'2002) (CD-ROM)
- Pawlak Z (1991) Rough sets. Kluwer, Dordrecht
- Skowron A, Grzymala-Busse J (1994) From rough set theory to evidence theory. In: Yager R, Fedrizzi M, Kacprzyk J (eds) *Advances in the Dempster–Shafer theory of evidence*. Wiley, New York, pp 193–236
- Tsumoto S (2000) Knowledge discovery in clinical databases and evaluation of discovered knowledge in out-patient clinic. *Inf Sci* 124:125–137
- Tsumoto S (2003) Statistical independence as linear independence. In: Skowron A, Szczuka M (eds) *Electronic notes in theoretical computer science*, vol 82. Elsevier
- Tsumoto S, Tanaka H (1996) Automated discovery of medical expert system rules from clinical databases based on rough sets. In: Proceedings of the second international conference on knowledge discovery and data mining 96. AAAI Press, Palo Alto, pp 63–69

Fuzzy Logic

Lotfi A. Zadeh

Department of EECS, University of California,
Berkeley, CA, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Conceptual Structure of Fuzzy Logic

The Basics of Fuzzy Set Theory

The Concept of Granulation

The Concepts of Precisation and Cointensive
Precisation

The Concept of a Generalized Constraint

Principal Contributions of Fuzzy Logic

A Glimpse of What Lies Beyond Fuzzy Logic

Bibliography

Glossary

Cointension A qualitative measure of proximity of meanings/input-output relations.

Extension principle A principle which relates to propagation of generalized constraints.

f-validity fuzzy validity.

Fuzzy if-then rule A rule of the form: if X is A then Y is B . In general, A and B are fuzzy sets.

Fuzzy logic (FL) A precise logic of imprecision, uncertainty and approximate reasoning.

Fuzzy logic gambit Exploitation of tolerance for imprecision through deliberate m-imprecisation followed by mm-precisation.

Fuzzy set A class with a fuzzy boundary.

Generalized constraint A constraint of the form X is r , where X is the constrained variable, R is the constraining relation and r is an indexical variable which defines the modality of the constraint, that is, its semantics. In general, generalized constraints have elasticity.

Generalized constraint language A language generated by combination and propagation of generalized constraints.

Graduation Association of a scale of degrees with a fuzzy set.

Granuland Result of granulation.

Granular variable A variable which takes granules as variables.

Granulation Partitioning of an object/set into granules.

Granule A clump of attribute values drawn together by indistinguishability, equivalence, similarity, proximity or functionality.

Linguistic variable A granular variable with linguistic labels of granular values.

m-precision Precision of meaning.

mh-precisiand m-precisiand which is described in a natural language (human-oriented).

mm-precisiand m-precisiand which is described in a mathematical language (machine-oriented).

p-validity provable validity.

Precisiand Result of precisation.

Preciend Object of precisation.

v-precision Precision of value.

Definition of the Subject

Viewed in a historical perspective, fuzzy logic is closely related to its precursor – fuzzy set theory (Zadeh 1965). Conceptually, fuzzy logic has a much broader scope and a much higher level of generality than traditional logical systems, among them the classical bivalent logic, multivalued logics, model logics, probabilistic logics, etc. The principal objective of fuzzy logic is formalization – and eventual mechanization – of two remarkable human capabilities. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. And second, the capability to perform a wide variety of physical and mental tasks – such as driving a car

in city traffic and summarizing a book – without any measurement and any computations.

A concept which has a position of centrality in fuzzy logic is that of a fuzzy set. Informally, a fuzzy set is a class with a fuzzy boundary, implying a gradual transition from membership to non-membership. A fuzzy set is precisiated through graduation, that is, through association with a scale of grades of membership. Thus, membership in a fuzzy set is a matter of degree. Importantly, in fuzzy logic everything is or is allowed to be graduated, that is, be a matter of degree. Furthermore, in fuzzy logic everything is or is allowed to be granulated, with a granule being a clump of attribute-values drawn together by indistinguishability, equivalence, similarity, proximity or functionality. Graduation and granulation form the core of fuzzy logic. Graduated granulation is the basis for the concept of a linguistic variable – a variable whose values are words rather than numbers (Zadeh 1973). The concept of a linguistic variable is employed in almost all applications of fuzzy logic.

During much of its early history, fuzzy logic was an object of controversy stemming in part from the pejorative connotation of the term “fuzzy”. In reality, fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and uncertainty.

An important milestone in the evolution of fuzzy logic was the development of the concept of a linguistic variable and the machinery of fuzzy if-then rules (Zadeh 1973, 1996). Another important milestone was the conception of possibility theory (Zadeh 1978a). Possibility theory and probability theory are complimentary. A further important milestone was the development of the formalism of computing with words (CW) (Zadeh 1999). Computing with words opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories.

In the following, fuzzy logic is viewed in a nontraditional perspective. In this perspective, the cornerstones of fuzzy logic are graduation, granulation, precisiation and the concept of a generalized constraint. The concept of a generalized constraint serves to precisiate the concept of granular information. Granular information is the

basis for granular computing (GrC) (Bargiela and Pedrycz 2002; Jankowski and Skowron 2007; Lin 1997; Zadeh 1979a, 1997, 1998). In granular computing the objects of computation are granular variables, with a granular value of a granular variable representing an imprecise and/or uncertain information about the value of the variable. In effect, granular computing is the computational facet of fuzzy logic. GrC and CW are closely related. In coming years, GrC and CW are likely to play increasingly important roles in the evolution of fuzzy logic and its applications.

Introduction

Science deals not with reality but with models of reality. In large measure, scientific progress is driven by a quest for better models of reality.

In the real world, imprecision, uncertainty and complexity have a pervasive presence. In this setting, construction of better models of reality requires a better understanding of how to deal effectively with imprecision, uncertainty and complexity. To a significant degree, development of fuzzy logic has been, and continues to be, motivated by this need.

In essence, logic is concerned with formalization of reasoning. Correspondently, fuzzy logic is concerned with formalization of fuzzy reasoning, with the understanding that precise reasoning is a special case of fuzzy reasoning.

Humans have many remarkable capabilities. Among them there are two that stand out in importance. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. And second, the capability to perform a wide variety of physical and mental tasks – such as driving a car in heavy city traffic and summarizing a book – without any measurements and any computations. In large measure, fuzzy logic is aimed at formalization, and eventual mechanization, of these capabilities. In this perspective, fuzzy logic plays the role of a bridge from natural to machine intelligence.

There are many misconceptions about fuzzy logic. A common misconception is that fuzzy logic is fuzzy. In reality, fuzzy logic is not fuzzy. Fuzzy logic deals precisely with imprecision and uncertainty. In fuzzy logic, the objects of deduction are, or are allowed to be fuzzy, but the rules governing deduction are precise. In summary, fuzzy logic is a precise system of reasoning, deduction and computation in which the objects of discourse and analysis are associated with information which is, or is allowed to be, imprecise, uncertain, incomplete, unreliable, partially true or partially possible. For illustration, here are a few simple examples of reasoning in which the objects of reasoning are fuzzy.

First, consider the familiar example of deduction in Aristotelian, bivalent logic.

$$\frac{\begin{array}{l} \text{all men are mortal} \\ \text{Socrates is a man} \end{array}}{\text{Socrates is mortal}}$$

In this example, there is no imprecision and no uncertainty. In an environment of imprecision and uncertainty, an analogous example – an example drawn from fuzzy logic – is

$$\frac{\begin{array}{l} \text{most Swedes are tall} \\ \text{Magnus is a Swede} \end{array}}{\text{it is likely that Magnus is tall}}$$

with the understanding that Magnus is an individual picked at random from a population of Swedes. To deduce the answer from the premises, it is necessary to precisiate the meaning of “most” and “tall,” with “likely” interpreted as a fuzzy probability which, as a fuzzy number, is equal to “most”. This simple example points to a basic characteristic of fuzzy logic, namely, in fuzzy logic precisiation of meaning is a prerequisite to deduction. In the example under consideration, deduction is contingent on precisiation of “most”, “tall” and “likely”. The issue of precisiation has a position of centrality in fuzzy logic.

In fuzzy logic, deduction is viewed as an instance of question-answering. Let I be an information set consisting of a system of propositions

$p_1, \dots, p_n$, $I = S(p_1, \dots, p_n)$. Usually, I is a conjunction of $p_1, \dots, p_n$. Let q be a question. A question-answering schema may be represented as

$$\frac{I}{\frac{q}{\text{ans}(q/I)}}$$

where $\text{ans}(q/I)$ denotes the answer to q given I . The following examples are instances of the deduction schema

$$\begin{array}{l} I: \text{ most Swedes are tall} \\ \text{Magnus is a Swede} \\ q: \text{ what is the probability that Magnus is tall?} \end{array} \quad (1)$$

$\text{ans}(q/I)$ is likely, likely=most

$$\begin{array}{l} I: \text{ most Swedes are tall} \\ q: \text{ what fraction of Swedes are not tall?} \end{array} \quad (2)$$

$\text{ans}(q/I)$ is $(1 - \text{most})$

$$\begin{array}{l} I: \text{ most Swedes are tall} \\ q: \text{ what fraction of Swedes are short?} \end{array} \quad (3)$$

$\text{ans}(q/I)$

$$\begin{array}{l} I: \text{ most Swedes are tall} \\ q: \text{ what is the average height of Swedes?} \end{array} \quad (4)$$

$\text{ans}(q/I)$

$$\begin{array}{l} I: \text{ a box contains balls of various sizes} \\ \text{most are small} \\ \text{there are many more small balls than large balls} \\ q: \text{ what is the probability that a ball} \\ \text{drawn at random is neither large nor small?} \end{array} \quad (5)$$

$\text{ans}(q/I)$.

In these examples, rules of deduction in fuzzy logic must be employed to compute $\text{ans}(q/I)$. For (1) and (2) deduction is simple. For (3)–(5) deduction requires the use of what is referred to as the extension principle (Zadeh 1965, 1975a). This principle is discussed in section “[The Concept of a Generalized Constraint](#)”.

A less simple example of deduction involves interpolation of an imprecisely specified function. Interpolation of imprecisely specified functions, or interpolative deduction for short, plays a pivotal role in many applications of fuzzy logic, especially in the realm of control.

For simplicity, assume that f is a function from reals to reals, $Y = f(X)$. Assume that what is known about f is a collection of input-output pairs of the form

$$*f = ((*a_1, *b_1), \dots, (*a_n, *b_n)),$$

where $*a$ is an abbreviation of “approximately a ”. Such a collection is referred to as a fuzzy graph of f (Zadeh 1974). A fuzzy graph of f may be interpreted as a summary of f . In many applications, a fuzzy graph is described as a collection of fuzzy if-then rules of the form

$$\text{if } X \text{ is } *a_i \text{ then } Y \text{ } *b_i, \quad i = 1, \dots, n.$$

Let a be a value of X . Viewing $*f$ as an information set, I , interpolation of f may be expressed as a question-answering schema

$$\frac{\begin{array}{l} I: *f \\ q: *f(*a) \end{array}}{\text{ans}(q(*f, *a))}$$

A very simple example of interpolative deduction is the following. Assume that $*a_1, *a_2, *a_3$ are labeled small, medium and large, respectively. A fuzzy graph of f may be expressed as a calculus of fuzzy if-then rules.

$$\begin{array}{l} *f: \text{ if } X \text{ is small then } Y \text{ is small} \\ \quad \text{if } X \text{ is medium then } Y \text{ is large} \\ \quad \text{if } X \text{ is large then } Y \text{ is small.} \end{array}$$

Given a value of X , $X = a$, the question is: What is $*f(*a)$? The so-called Mamdani rule (Mamdani and Assilian 1975; Zadeh 1974, 1976) provides an answer to this question.

More generally, in an environment of imprecision and uncertainty, fuzzy if-then rules may be of the form

if X is $*a_i$, then usually (Y is $*b_i$), $i = 1, \dots, n$.

Rules of this form endow fuzzy logic with the capability to model complex causal dependencies, especially in the realms of economics, social systems, forecasting and medicine.

What is not widely recognized is that fuzzy logic is more than an addition to the methods of dealing with imprecision, uncertainty and complexity. In effect, fuzzy logic represents a paradigm shift. More specifically, it is traditional to associate scientific progress with progression from perceptions to numbers. What fuzzy logic adds to this capability are four basic capabilities.

- (a) Nontraditional. Progression from perceptions to precisiated words.
- (b) Nontraditional. Progression from unprecisiated words to precisiated words.
- (c) Countertraditional. Progression from numbers to precisiated words.
- (d) Nontraditional. Computing with words (CW)/NL-computation.

These capabilities open the door to a wide-ranging enlargement of the role of natural languages in scientific theories.

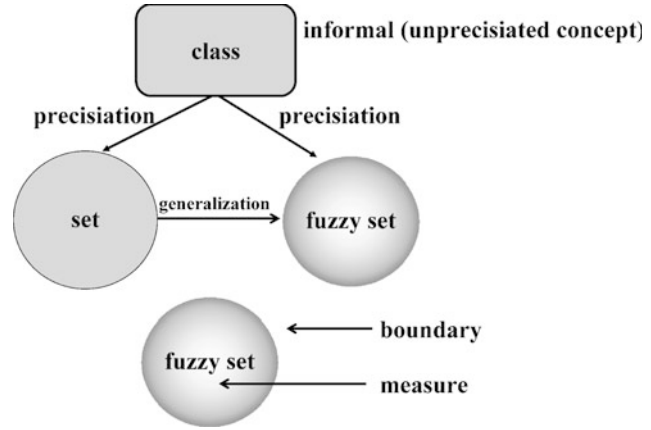
Our brief discussion of deduction in the context of fuzzy logic is intended to clarify the nature of problems which fuzzy logic is designed to address. The principal concepts and techniques which form the core of fuzzy logic are discussed in the following. Our discussion draws on the concepts and ideas introduced in (Zadeh 2008).

Conceptual Structure of Fuzzy Logic

There are many logical systems, among them the classical, Aristotelian, bivalent logic, multivalued logics, model logics, probabilistic logic, logic, dynamic logic, etc. What differentiates fuzzy logic from such logical systems is that fuzzy logic is much more than a logical system.

The point of departure in fuzzy logic is the concept of a fuzzy set. Informally, a fuzzy set is a class with a fuzzy boundary, implying that, in

Fuzzy Logic, Fig. 1 The concepts of a set and a fuzzy set are derived from the concept of a class through precisiation. A fuzzy set has a fuzzy boundary. A fuzzy set is precisiated through graduation



general, transition from membership to non-membership in a fuzzy set is gradual rather than abrupt. A set is a class with a crisp boundary (Fig. 1). A set, A , in a space U , $U = \{u\}$, is precisiated through association with a characteristic function which maps U into $\{0, 1\}$. More generally, a fuzzy set, A , is precisiated through graduation, that is, through association with A of a membership function, μ_A – a mapping from U to a grade of membership space, G , with $\mu_A(u)$ representing the grade of membership of u in A . In other words, membership in a fuzzy set is a matter of degree. A familiar example of graduation is the association of Richter scale with the class of earthquakes. A fuzzy set is basic if G is the unit interval. More generally, G may be a partially ordered set. L-fuzzy sets (Goguen 1967) fall into this category. A basic fuzzy set is of Type 1. A fuzzy set, A , is of Type 2 if $\mu_A(u)$ is a fuzzy set of Type 1. Recursively, a fuzzy set, A , is of Type n if $\mu_A(u)$ is a fuzzy set of Type $n - 1$, $n = 2, 3, \dots$ (Zadeh 1975a). Fuzzy sets of Type 2 have become an object of growing attention in the literature of fuzzy logic (Mendel 2001). Unless stated to the contrary, a fuzzy set is assumed to be of Type 1 (basic).

Note. A clarification is in order. Consider a concatenation of two words, A and B , with A modifying B , e. g. A is an adjective and B is a noun. Usually, A plays the role of an s-modifier, that is, a modifier which specializes B in the sense that AB is a subset of B , as in convex set. In some instances, however, A plays the role of a g-modifier, that is, a modifier which generalizes

B . In this sense, fuzzy in fuzzy set, fuzzy logic and, more generally, in fuzzy B , is a g-modifier. Examples: fuzzy topology, fuzzy measure, fuzzy arithmetic, fuzzy stability, etc. Many misconceptions about fuzzy logic are rooted in incorrect interpretation of fuzzy as an s-modifier.

What is important to note is that (a) $\mu_A(u)$ is name-based (extensional) if u is the name of an object in U , e. g., $\mu_{\text{middle-aged}}$ (Vera); (b) $\mu_A(u)$ is attribute-based (intensional) if the grade of membership is a function of an attribute of u , e. g., age; and (c) $\mu_A(u)$ is perception-based if u is a perception of an object, e. g., Vera's grade of membership in the class of middle-aged women, based on her appearance, is 0.8.

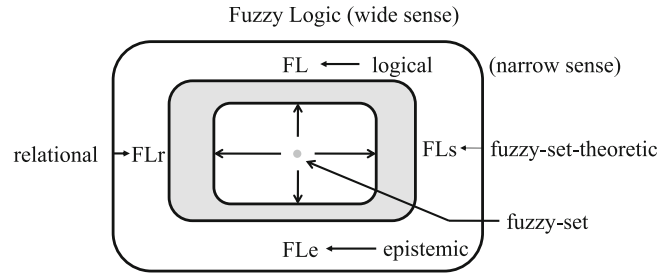
It should be observed that a class of objects, A , has two basic attributes: (a) the boundary of A ; and (b) the cardinality (count) or, more generally, the measure of A (Fig. 1). In this perspective, fuzzy set theory is, in the main, boundary-oriented, while probability theory is, in the main, measure-oriented. Fuzzy logic is, in the main, both boundary- and measure-oriented. The concept of a membership function is the centerpiece of fuzzy set theory.

With the concept of a fuzzy set as the point of departure, different directions may be pursued, leading to various facets of fuzzy logic. More specifically, the following are the principal facets of fuzzy logic: the logical facet, FLI; the fuzzy-set-theoretic facet, FLs, the epistemic facet, Fle; and the relational facet, FLr (Fig. 2).

The logical facet of FL, FLI, is fuzzy logic in its narrow sense. FLI may be viewed as a

Fuzzy Logic,

Fig. 2 Principal facets of fuzzy logic (FL). The nucleus of fuzzy logic is the concept of a fuzzy set



Fuzzy Logic, Fig. 3 The cornerstones of a nontraditional view of fuzzy logic



generalization of multivalued logic. The agenda of FLI is similar in spirit to the agenda of classical logic (Esteva and Godo 2007; Godo et al. 1991; Hajek 1998; Novak et al. 1999; Novak 2006).

The fuzzy-set-theoretic facet, FLs, is focused on fuzzy sets. The theory of fuzzy sets is central to fuzzy logic. Historically, the theory of fuzzy sets (Zadeh 1965) preceded fuzzy logic (Zadeh 1975c). The theory of fuzzy sets may be viewed as an entry to generalizations of various branches of mathematics, among them fuzzy topology, fuzzy measure theory, fuzzy graph theory, fuzzy algebra and fuzzy differential equations. Note that fuzzy X is a fuzzy-set-theory-based or, more generally, fuzzy-logic-based generalization of X .

The epistemic facet of FL, FLe, is concerned with knowledge representation, semantics of natural languages and information analysis. In FLe, a natural language is viewed as a system for describing perceptions. An important branch of FLe is possibility theory (Dubois and Prade 1988; Zadeh 1978a, 1981). Another important branch of FLe is the computational theory of perceptions (Zadeh 1999, 2000, 2001).

The relational facet, FLr, is focused on fuzzy relations and, more generally, on fuzzy dependencies. The concept of a linguistic variable – and the associated calculi of fuzzy if-then rules – play pivotal roles in almost all applications of fuzzy logic (Aliev et al. 2006; Bardossy and Duckstein 1995; Bezdek and Pal 1992; Bouchon-Meunier

et al. 2000; Driankov et al. 1993; Dubois and Prade 1980; Filev and Yager 1994; Hirota and Sugeno 1995; Jamshidi et al. 1997a; Kandel and Langholz 1994; Lawry et al. 2003; Mendel 2001; Pedrycz and Gomide 2007; Ross 2004; Yager and Zadeh 1992; Yen et al. 1995; Yen and Langari 1998; Ying 2000).

The cornerstones of fuzzy logic are the concepts of graduation, granulation, precisiation and generalized constraints (Fig. 3). These concepts are discussed in the following.

The Basics of Fuzzy Set Theory

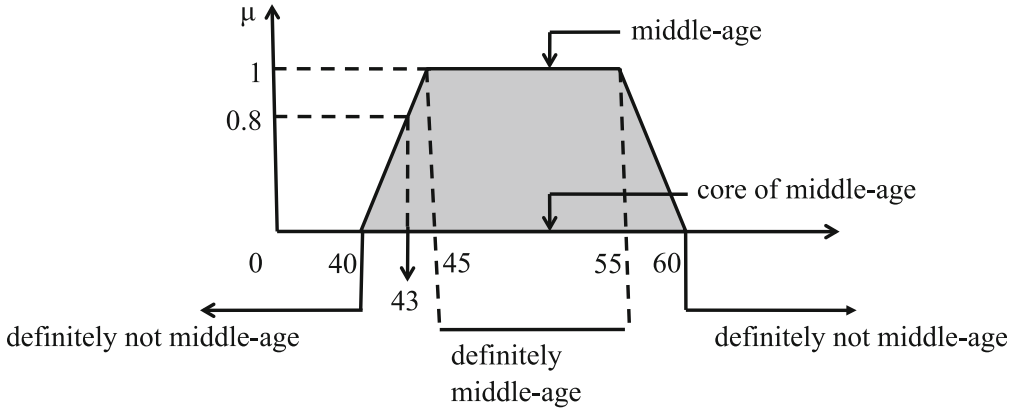
The grade of membership, $\mu_A(u)$, of u in A may be interpreted in various ways. It is convenient to employ the proposition p : Vera is middle-aged, as an example, with middle-age represented as a fuzzy set shown in Fig. 4. Among the various interpretations of p are the following.

Assume that q is the proposition: Vera is 43 years old; and r is the proposition: the grade of membership of Vera in the fuzzy set of middle-aged women is 0.8.

- The truth value of p given r is 0.8.
- The possibility of q given p and r is 0.8.
- The degree to which the concept of middle-age has to be stretched to apply to Vera is $(1-0.8)$.

- Imprecision of meaning = elasticity of meaning
- Elasticity of meaning = fuzziness of meaning

Example: middle age



Fuzzy Logic, Fig. 4 Precisation of middle-age through graduation

- (d) 80% of voters in a voting group vote for p given q and r .
 (e) The probability of p given q and r is 0.8.

$$\int_U \mu_A(u)/u.$$

Of these interpretations, (a)–(c) are most compatible with intuition.

If A and B are fuzzy sets in U , then their intersection (conjunction), $A \cap B$, and union (disjunction), $A \cup B$, are defined as

$$\begin{aligned}\mu_{A \cap B}(u) &= \mu_A(u) \wedge \mu_B(u), \quad u \in U \\ \mu_{A \cup B}(u) &= \mu_A(u) \vee \mu_B(u), \quad u \in U\end{aligned}$$

where $\wedge = \min$ and $\vee = \max$. More generally, conjunction and disjunction are defined through the concepts of t-norm and t-conorm, respectively (Pedrycz and Gomide 2007).

When U is a finite set, $U = \{u_1, \dots, u_n\}$, it is convenient to represent A as a union of fuzzy singletons,

$$\begin{aligned}\mu_A(u_i)/u_i, \quad i = 1, \dots, n. \\ \text{Specifically, } A = \mu_A(u_1)/u_1 + \dots + \mu_A(u_n)/u_n\end{aligned}$$

in which $+$ denotes disjunction. More compactly,

$$A = \sum_i \mu_A(u_i)/u_i, \quad i = 1, \dots, n.$$

When U is a continuum, A may be expressed as

A basic concept in fuzzy set theory is that of a level set (Zadeh 1965), commonly referred to as an α -cut (Pedrycz and Gomide 2007). Specifically, if A is a fuzzy set in U , then an α -cut, A_α , is defined as (Fig. 5)

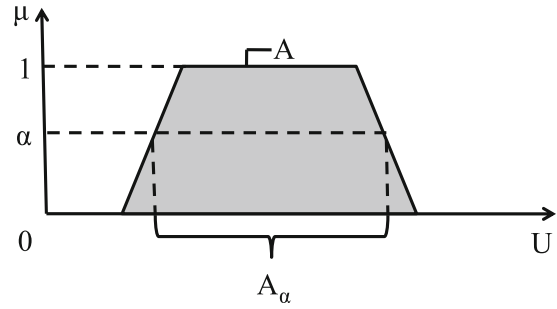
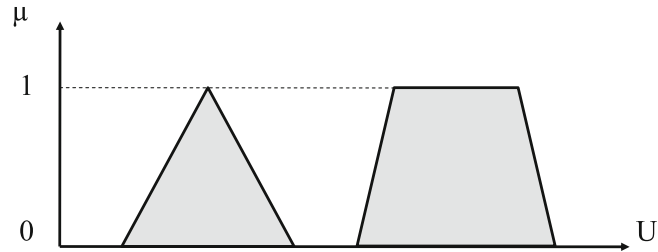
$$A_\alpha = \{u | \mu_A(u) \geq \alpha\}, \quad 0 < \alpha \leq 1.$$

The core of A is the α -cut with $\alpha = 1$.

A fuzzy set, A , is a fuzzy interval if all of its α -cuts are intervals. A fuzzy set, A , is convex if all of its α -cuts are convex (Zadeh 1965).

In most practical applications, a membership function is assumed to have a simple form, e. g. a triangle or a trapezoid (Fig. 6). For convenience, triangular and trapezoidal fuzzy sets are abbreviated as tr-sets and tp-sets, respectively. It should be noted that a singleton is associated with one parameter, an interval with two parameters, a tr-set with three parameters and a tp-set with four parameters. The number of parameters may be interpreted as the number of degrees of freedom. In general, the number of degrees of freedom is covariant with closeness of approximation.

A concept which plays an important role in fuzzy set theory is that of cardinality, that is, the

Fuzzy Logic,**Fig. 5** Definition of α -cut**Fuzzy Logic,****Fig. 6** Triangular and trapezoidal fuzzy sets. tr-sets and tp-sets

count of elements in a fuzzy set (Hajek 1998; Ralescu 1995; Wygralak 2003; Zadeh 1983b). Basically, there are two ways in which cardinality can be defined: (a) crisp (scaler) cardinality and (b) fuzzy (stratified) cardinality. In the case of (a), the count of elements in a fuzzy set is a crisp number; in the case of (b) it is a fuzzy set.

More specifically, consider a fuzzy set, A , defined in $U = \{u_1, \dots, u_n\}$ through its membership function, $\mu_A(u)$. The sigma-count of A is defined as

$$\Sigma \text{Count}(A) = \sum_{i=1}^n \mu_A(u_i).$$

If A and B are fuzzy sets defined in U , then the relative sigma-count, $\Sigma \text{Count}(A/B)$, is defined as

$$\begin{aligned} \Sigma \text{Count}(A/B) &= \frac{\Sigma \text{Count}(A \cap B)}{\Sigma \text{Count}(B)} \\ &= \frac{\sum_{i=1}^n \mu_A(u_i) \wedge \mu_B(u_i)}{\sum_{i=1}^n \mu_B(u_i)}, \end{aligned}$$

where $\wedge = \min$, and summations are arithmetic.

A stratified count, $S \text{Count}(A)$, is a fuzzy set. As such, it is more informative than $\Sigma \text{Count}(A)$ but has the disadvantage of greater complexity.

A stratified count is defined in terms of α -cuts of A (Zadeh 1983b). Specifically, let $A_{\alpha_1}, \dots, A_{\alpha_n}$ be α -cuts of A , with $\alpha_1 < \alpha_2 < \dots < \alpha_n \leq 1$. Then

$$S \text{Count}(A) = \alpha_1 / \text{Count}(A_{\alpha_2}) + \dots + \alpha_n / \text{Count}(A_{\alpha_n})$$

or equivalently,

$$S \text{Count}(A) = \{(\alpha_i, A_{\alpha_i})\}, \quad i = 1, \dots, n.$$

As a simple illustration, consider the fuzzy set

$$A = 0.4/u_1 + 0.8/u_2 + 1/u_3 + 0.9/u_4 + 0.3/u_5.$$

In this case,

$$A_{0.3} = \{u_1, u_2, u_3, u_4, u_5\}, \quad \text{Count}(A_{0.3}) = 5$$

$$A_{0.4} = \{u_1, u_2, u_3, u_4\}, \quad \text{Count}(A_{0.4}) = 4$$

$$A_{0.8} = \{u_2, u_3, u_4\}, \quad \text{Count}(A_{0.8}) = 3$$

$$A_{0.9} = \{u_3, u_4\}, \quad \text{Count}(A_{0.9}) = 2$$

$$A_1 = \{u_3\}, \quad \text{Count}(A_1) = 1$$

$$S \text{Count}(A) = 1/1 + 0.9/2 + 0.8/3 + 0.4/4 + 0.3/5$$

$$\Sigma \text{Count}(A) = 3.4.$$

The concept of a membership function has a position of centrality in fuzzy logic. In this context, a natural question is: How can a membership function be constructed? There are three basic

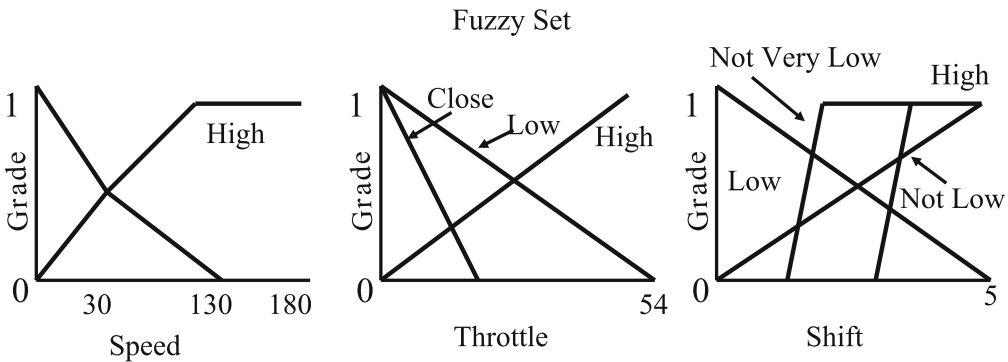
approaches: graduation through declaration; graduation through composition/deduction; and graduation through exemplification/ostention/elicitation.

Graduation through declaration is employed in many applications of fuzzy logic, especially in the realm of control systems. In this mode of graduation, a membership function is declared (specified) by the designer of a system. An example is the Honda fuzzy logic transmission (Fig. 7).

Declarative graduation has the potential for important applications in the realm of precision/standardization of many concepts and terms in various fields of science, especially in

the realms of medicine and economics. A very simple example is standardization of the meaning of normal temperature, mild fever, high fever, etc. It is convenient to standardize the meaning of such terms through the use of trapezoidal membership functions, that is, tp-sets (Fig. 8).

In many cases, a membership function is associated with a number of adjustable parameters, e. g. the four parameters which define a tp-set. The parameters are adjusted through experimentation or self-learning. For this purpose, techniques drawn from neurocomputing and evolutionary computing are commonly employed (Rutkowska 2002; Rutkowski 2008).

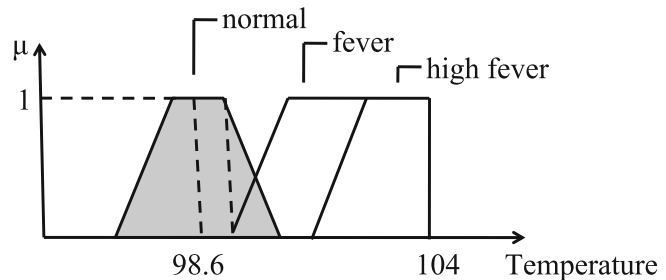


Control fuzzy if-then rules:

1. If (speed is low) and (shift is high) then (-3)
2. If (speed is high) and (shift is low) then (+3)
3. If (throt is low) and (speed is high) then (+3)
4. If (throt is low) and (speed is low) then (+1)
5. If (throt is high) and (speed is high) then (-1)
6. If (throt is high) and (speed is low) then (-3)

Fuzzy Logic, Fig. 7 Declared membership functions and associated fuzzy if-then rules in Honda's fuzzy logic transmission

Fuzzy Logic, Fig. 8 Declarative graduation/standardization of normal temperature, fever and high fever



In construction through composition/deduction, a membership function is composed from other membership functions. Composition/deduction may be simple e. g. a combination of conjunctive or disjunctive operations. More generally, composition/deduction is associated with a chain of deductions which involve generalized constraint propagation. Propagation of generalized constraints is discussed in section “[The Concept of a Generalized Constraint](#)”.

A special case of composition which plays an important role in fuzzy set theory involves the concept of a convex combination (Zadeh 1965). More specifically, let $A_1, \dots, A_n$ be a collection of crisp or fuzzy sets in U . Let $a_1, \dots, a_n$ be numbers in $[0,1]$ which add up to unity. A fuzzy set, A , is a convex combination of the A_i , $i = 1, \dots, n$, if

$$A = a_1A_1 + \dots + a_nA_n,$$

with the understanding that

$$\mu_A(u) = a_1\mu_{A_1}(u) + \dots + a_n\mu_{A_n}(u).$$

What is important to note is that a convex combination of crisp sets is a fuzzy set. If the coefficients $a_1, \dots, a_n$ are interpreted as probabilities, then A may be interpreted as the expected value of a random set. This connection between fuzzy sets and random sets has been an object of considerable attention in the literature of fuzzy logic (Goodman and Nguyen 1985; Ogura et al. 2002; Orlov 1980).

Underlying graduation through exemplification is the remarkable human capability to rank-order perceptions. It should be noted that humans learn the meaning of terms and concepts mostly through exemplification.

It is convenient to employ an example to describe graduation through exemplification. Assume that A tells B that Vera is middle-aged. B can elicit A 's meaning of middle-aged by asking A to mark on the scale $[0, 1]$ the degree to which a particular age, say 43, fits A 's meaning of middle-aged. The process is repeated for various values of age. Eventually, the collected data are employed to approximate to A 's meaning of middle-aged by a trapezoidal membership function. In the case of fuzzy sets of Type 2, the mark is a fuzzy point.

A basic feature of fuzzy logic is: In fuzzy logic everything is or is allowed to be graduated.

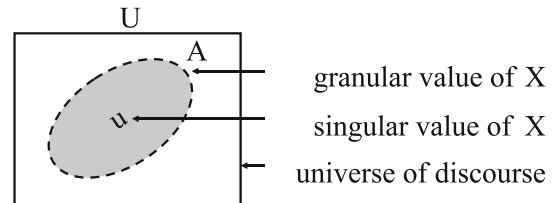
The Concept of Granulation

The concept of granulation is unique to fuzzy logic and plays a pivotal role in its applications (Zadeh 1979a, 1997). The concept of granulation is inspired by the way in which humans deal with imprecision, uncertainty and complexity.

Granulation is rooted in the concept of a granule. Informally, a granule, G , in a universe of discourse, U , is a clump of elements of U which are drawn together by indistinguishability, equivalence, similarity, proximity or functionality (Fig. 9). For example, an interval is a granule.

Fuzzy Logic,

Fig. 9 Singular and granular values



	singular	granular
unemployment	7.3%	high
temperature	102.5	very high
blood pressure	160/80	high

So is a Gaussian distribution, and so is a fuzzy interval. A granule, G , is precisiated through association of G with a generalized constraint – a concept which will be defined in section “[The Concept of a Generalized Constraint](#)”. The concept of a generalized constraint is more general than that of a membership function.

A singular variable takes singletons in U as values. A granular variable, X , is a variable which takes granules as values. For example, age is a granular variable if its values are young, middle-aged and old. A linguistic variable is a granular variable whose granular values carry linguistic labels (Fig. 10). The concept of a linguistic variable (Zadeh 1973, 1975a) is employed in almost all applications of fuzzy logic.

Granulation is an operation which maps singletons in U into granules. Granulation applied to a singular variable, X , transforms X into a granular variable, $*X$. For example, if age is a real-valued variable taking values in the interval $[0, 120]$, granulation of X transforms X into a linguistic variable, $*X$, with values young, middle-aged and old. When convenient, the result of granulation is referred to as the granuland.

Granulation of a set, A , results in a partition of A into granules. For example, granulation applied to the interval $[0, 120]$ results in the granules young, middle-aged and old.

The membership function of a fuzzy set takes values in the interval $[0, 1]$. Not infrequently, the grade of membership is not known precisely. In this case, it may be expedient to granulate the interval $[0, 1]$, with the granular values of membership being zero, low, medium, high and

1. Such granular values of membership are particularly useful in dealing with fuzzy sets of Type 2.

Granulation may be applied to arbitrarily complex objects. Application of granulation to an expression involves replacement of one or more singular variables in the expression with granular variables. For example, if

$$Z = X + Y$$

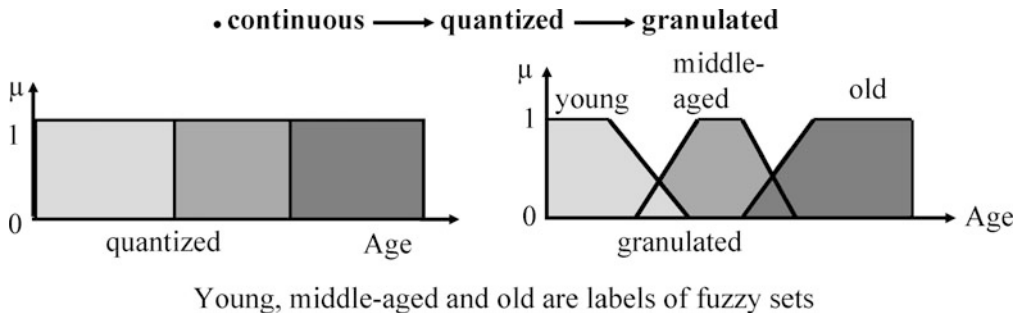
is an arithmetic expression, then its granuland may be expressed as

$$*Z = *X + *Y$$

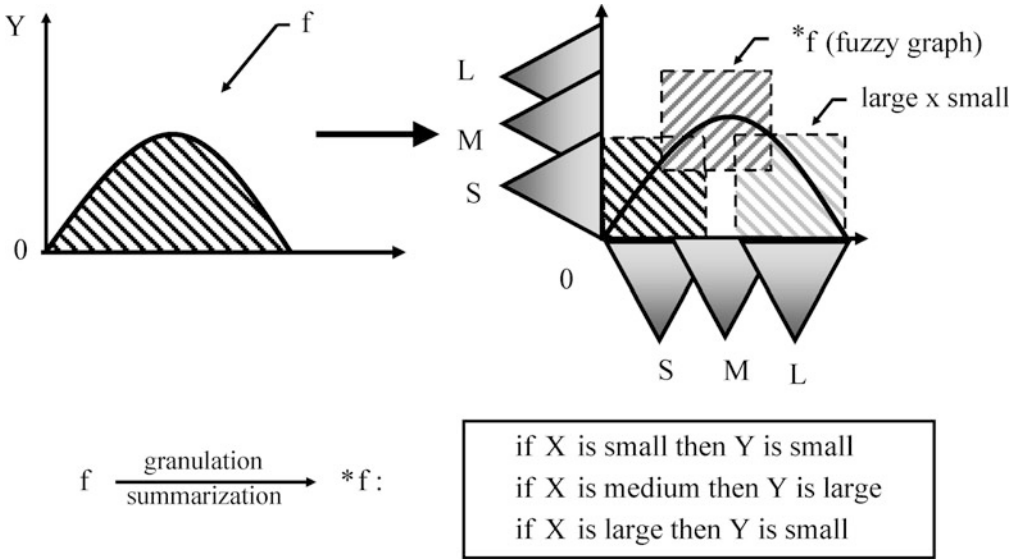
in which the starred variables are granular variables. In this sense, interval arithmetic may be viewed as the result of granulation of numerical arithmetic. Figure 11 shows an application of granulation to a function, f . The result of granulation, $*f$, is the fuzzy graph of f (Zadeh 1974, 1976). The fuzzy graph of f may be described as a collection of fuzzy if-then rules; $*f$ may be viewed as a summary of f . Describing a function as a collection of fuzzy if-then rules may be regarded as a form of information compression.

Application of granulation to probability distributions is illustrated in Fig. 12. The granules of X play the role of fuzzy events and the P_i are their granular probabilities (Zadeh 2005b).

It should be pointed out that in the case of probability distributions it is necessary to differentiate between granular probability distributions and granule-valued probability distributions (Zadeh 2002, 2006b) (Fig. 13). It should be noted that there is a connection between the

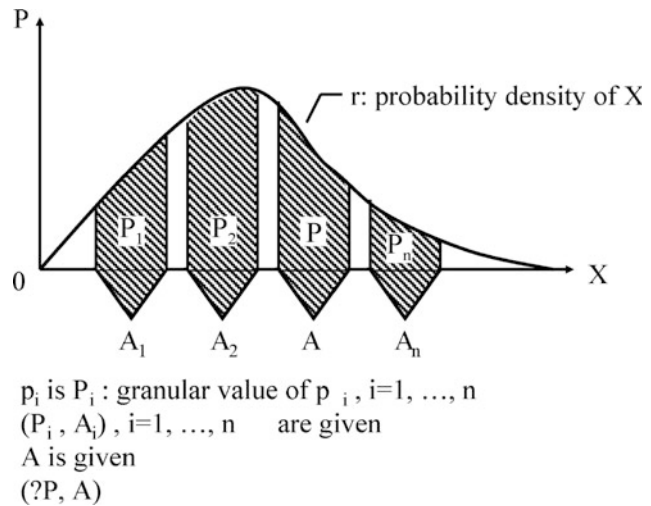


Fuzzy Logic, Fig. 10 Granulation of age; young, middle-aged and old are linguistic (granular) values of age



Fuzzy Logic, Fig. 11 Granulation of a function. S (small), M (medium) and L (large) are fuzzy sets. The granulation of f , $*f$, may be viewed as a summary of f

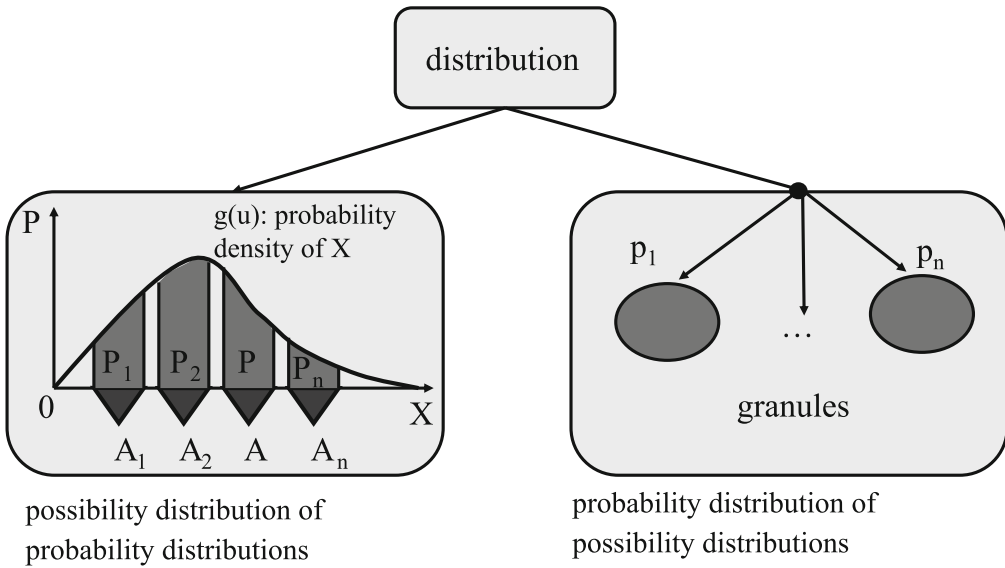
Fuzzy Logic, Fig. 12 Granulation and interpolation of a probability distribution



concept of a granular probability distribution and the notion of Perfilieva transform (Perfilieva 2007).

An instance of a granule-valued distribution is a random set. There is a close connection between granule-valued distributions and the Dempster–Shafer theory of evidence (Dempster 1967; Schum 1994; Shafer 1976). More about granular and granule-valued distributions will be said in section “The Concept of a Generalized Constraint”.

In addition to the concepts of graduation and granulation there are two basic concepts which play important roles in fuzzy logic. These are the concepts of precisiation (Zadeh 1984, 2004) and generalized constraint (Zadeh 1986b, 1999). These concepts along with the concepts of graduation and granulation form the cornerstones of fuzzy logic. The concepts of precisiation and cointensive precisiation are discussed in the following section. The concept of a generalized



Fuzzy Logic, Fig. 13 Granular vs. granule-valued distributions

constraint is discussed in section “[The Concept of a Generalized Constraint](#)”.

The Concepts of Precision and Cointensive Precision

There are not many concepts in science that are as pervasive as the concept of precision. There is an enormous literature. And yet, there are some important facets of the concept of precision which have received little if any attention. In particular, an issue that appears to have been overlooked relates to the need for differentiation between two forms of precision: precision of value, v-precision, and precision of meaning, m-precision (Zadeh 2006b). Specifically, consider a variable, X , whose value is not known precisely. In this event, the proposition $a \leq X \leq b$, where a and b are precisely specified numbers, is v-imprecise and m-precise. Similarly, the proposition: X is a normally distributed random variable with mean a and variance b , is v-imprecise and m-precise. On the other hand, the proposition: X is small, is both v-imprecise and m-imprecise if small is not defined precisely. If small is defined

precisely as a fuzzy set, then the proposition in question is v-imprecise and m-precise.

Informally, precision is an operation which transforms an object, p , into another object, p^* , which is more precisely defined, in some specified sense, than p . The reverse applies to imprecision. In the realm of our discourse p is a proposition, predicate, question, command or, more generally, a linguistic expression which has a semantic identity. Unless stated to the contrary, p will be assumed to be a proposition. For convenience, the object and the result of precision are referred to as precisend and precisand, respectively (Fig. 14).

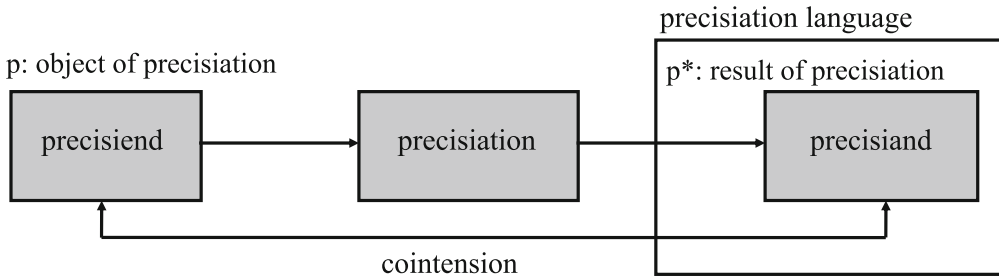
As in the case of precision/imprecision, there is a need for differentiation between v-precision/imprecision and m-precision/imprecision. Example:

$$X = 5 \xrightarrow[\text{m-imprecision}]{\text{v-imprecision}} X \text{ is small}$$

$$X \text{ is small} \xrightarrow{\text{m-precision}} X \text{ is small}$$

(small is defined as a fuzzy set).

It should be noted that data compression and summarization are forms of v-imprecision.



precisiand = model of meaning

cointension = measure of closeness of meanings of p and p^*

A precisiend may have many precisands.

precision = translation into precision language

Fuzzy Logic, Fig. 14 Basic concepts relating to precision and cointension

In the case of m-precision, there is a need for additional differentiation between m-precision which is human-oriented (non-mathematical), or mh-precision for short, and m-precision which is machine-oriented (mathematical), or mm-precision for short (Fig. 15). In this sense, a dictionary definition is a form of mh-precision, with the definiens and the definendum playing the roles of precisiend and precisiand, respectively. A mathematical definition of a concept, say stability, is an instance of mm-precision.

A convenient illustration of mh-precision and mm-precision is provided by the concept of “bear market”.

mh-precisiand	declining stock market with expectation of further decline
mm-precisiand	30 percent decline after 50 days, or a 13 percent decline after 145 days (Robert Shuster).

It is of interest to note that disambiguation is a form of m-precision. As an illustration, alternative interpretations of the predicate, P : Most tall Swedes, are shown in Fig. 16.

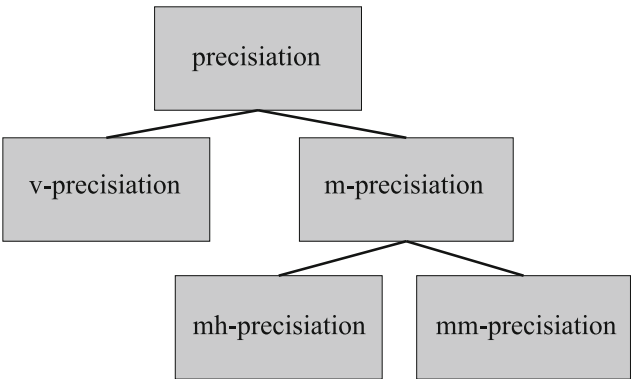
So far as v-imprecision is concerned, there is a need for differentiation between (a) v-imprecision which is forced; and (b) v-imprecision which is deliberate. To illustrate,

if the value of X is described imprecisely because the precise value of X is not known, then what is involved is forced v-imprecision. On the other hand, if the value of X is described imprecisely even though the precise value of X is known, then v-imprecision is deliberate.

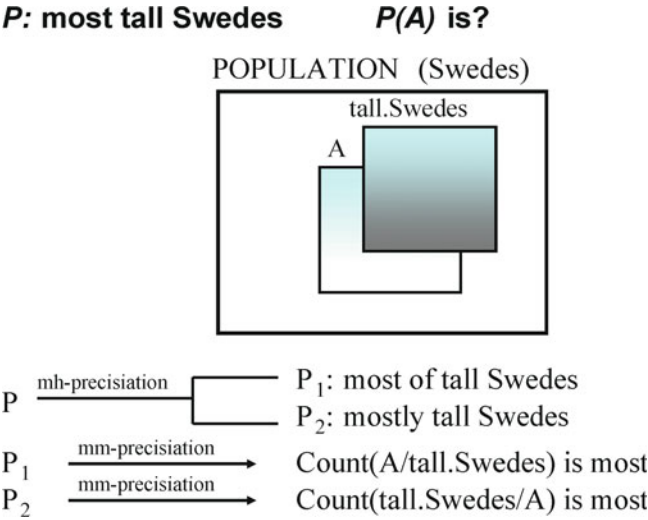
What is the point of deliberate v-imprecision? The rationale is that in many cases precision carries a cost. In such cases, deliberate v-imprecision serves a useful purpose if it provides a way of reducing cost. Familiar examples of deliberate v-imprecision are data compression and summarization. As will be seen at a later point, deliberate v-imprecision underlies the fuzzy logic gambit (section “[The Concept of a Generalized Constraint](#)”). The fuzzy logic gambit plays an important role in many applications of fuzzy logic, especially in the realm of consumer products – a realm in which cost is an important parameter.

A precisiend may be precisiated in a large, perhaps unbounded, number of ways. As an illustration, Fig. 17a and b shows some of the simpler mm-precisiands of the predicate “approximately a ,” or $*a$ for short. Note that the simplest mm-precisiand is a . This very simple mode of mm-precision is widely employed in many fields of science. Probability theory is a case in point. Most real-world probabilities are based on perceptions rather than on measurements.

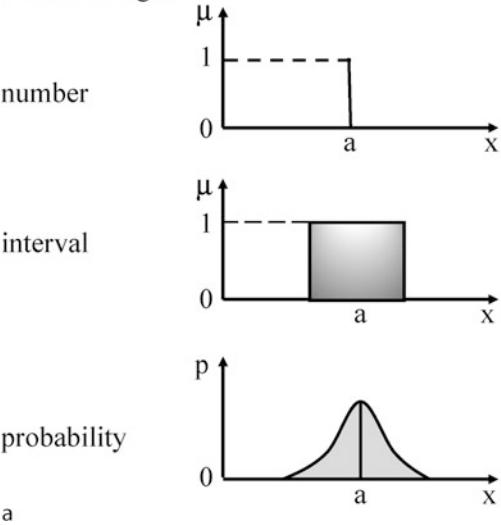
Fuzzy Logic,
Fig. 15 Modalities of
precisiation.
mh-precisiation =
nonmathematical
precisiation of meaning;
mm-precisiation =
mathematical precisiation
of meaning



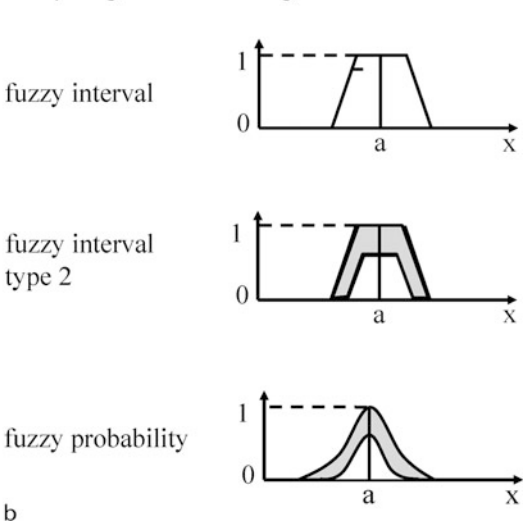
Fuzzy Logic,
Fig. 16 Disambiguation of
P: most tall Swedes



Bivalent Logic



Fuzzy Logic: Bivalent Logic + ...



Fuzzy Logic, Fig. 17 (a) Alternative modes of mm-precisiation of “approximately *a*,” **a*, within the framework of bivalent logic; (b) Alternative modes of mm-precisiation of “approximately *a*,” **a*, within the framework of fuzzy logic

Perceptions are intrinsically imprecise. As a consequence, so are most real-world probabilities. And yet, in most computations involving probabilities, probabilities are treated as exact numbers. For example, $^*0.7$ is treated as 0.7000.

The Concept of Cointension

Let p^* be a precisiand of p . It is expedient to associate with p^* two basic metrics: (a) cointension – a qualitative measure of the proximity of the meanings of p and p^* ; and (b) the computational complexity of p^* . In general, cointension and computational complexity are covariant in the sense that an increase in the cointension of p^* is associated with an increase in the computational complexity of p^* .

Cointension is a new term which is in need of clarification (Zadeh 2006b). In logic, intension and extension are defined, respectively, as attribute-based meaning, or i-meaning for short, and name-based meaning, or e-meaning for short (Belohlavek and Vychodil 2006; Cresswell 1973; Lambert and Van Fraassen 1970). For our purposes, it is helpful to interpret attribute-based as measurement-based and, name-based as perception-based. What this means is that a precisiend, p , is viewed as a perception of a concept, while its mm-precisiand, p^* , is viewed as a measurement-based definition of p . For example, in the case of the concept of bear market, we have.

mm-precisiand	30 percent decline after 50 days, or a 13 percent decline after 145 days (Robert Shuster).
----------------------	--

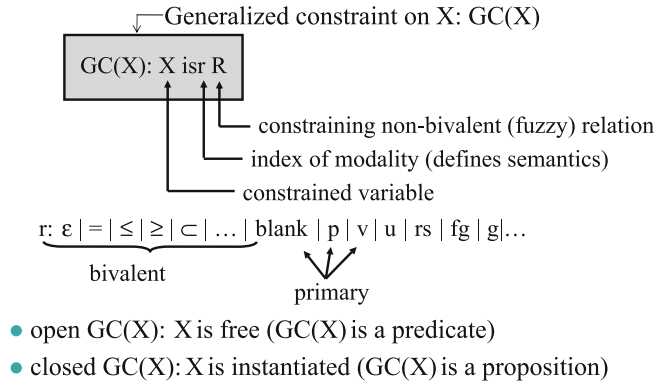
More concretely, if p^* is an mm-precisiand of p , then the cointension of p^* in relation to p , $C(p^*, p)$, is a qualitative measure of the degree of proximity of the i-meanings of p^* and p , or the proximity of the i-meaning of p^* and the e-meaning of p . p^* is cointensive if the degree of proximity is high. In the case of the bear market example, cointension is a measure of the degree to which the mm-precisiand fits our perception of “bear market”. Although this degree cannot be assessed precisely, it is evident that the degree is not high.

It should be noted that there is a close analogy between the concept of mm-precision and mathematical modeling. More specifically, the analog of mm-precision is mathematical modeling; the analog of precisiend is the object of modeling; the analog of precisiand is the model; and the analog of meaning is the input/output relation.

In the context of modeling, cointension is a measure of proximity of the input/output relations of the object of modeling and the model. A model is cointensive if its proximity is high.

The concept of cointension highlights an important issue. Specifically, what should be noted is that, in general, mm-precision of p is not the final objective. What matters is the cointension of an mm-precisiand of p . In general, what are sought are mm-precisiands which have high cointension. In other words, the desideratum is not merely mm-precision but, more importantly, cointensive mm-precision. As will be seen at a later point, a striking feature of fuzzy logic is its high power of cointensive precision.

The concept of cointensive mm-precision has an important implication for scientific theories. In large measure, scientific theories are based on bivalent logic. As a consequence, definitions of concepts are, generally, bivalent, in the sense that no degrees of truth are allowed. An example is the definition of bear market. The same applies to the definitions of recession, stability, independence, causality, stationarity, etc. The problem is that most of the concepts which are associated with bivalent definitions are in fact fuzzy, that is, are a matter of degree. For instance, the reason why the mm-precisiand of bear market is not cointensive is rooted in the fact that bear market is a fuzzy concept. What can be said in a general way is that bivalent-logic-based definitions of fuzzy concepts cannot be expected to be cointensive, just as linear models of nonlinear systems cannot be expected to be good models. In summary, an important conclusion relating to the concept of cointension may be stated as the Cointension Principle: To Achieve high cointension it is necessary, in general, to associate a fuzzy precisiend with a fuzzy precisiand.

Fuzzy Logic,**Fig. 18** Principal features of a generalized constraint

Note. In the sequel, unless stated to the contrary, precision should be understood as cointensive mm-precision

In the foregoing discussion, we talked about mm-precision but have not addressed a basic question: How can a proposition, p , be mm-precised? In fuzzy logic, a concept which plays a pivotal role in mm-precision is that of a generalized constraint. A brief discussion of this concept is presented in the following section.

The Concept of a Generalized Constraint

The concept of a constraint is high on the list of basic concepts in science. There is an extensive literature in the realms of mathematical programming and optimal control. Particularly worthy of note is the rapid growth of interest in constraint programming within computer science and related fields (Rossi and Codognet 2003).

A basic assumption which is commonly made in the literature is that constraints are hard (inelastic) and are precisely defined. This assumption is not a good fit to reality. In most realistic settings, constraints have some elasticity and are not precisely defined. Familiar examples are:

- Check-out time is 1 pm. A constraint on check-out time.
- Speed limit is 100 km/h. A constraint on speed.
- Vera has a son in mid-twenties and a daughter in mid-thirties. A constraint on Vera's age.

A first step toward precisiation of such constraints was taken in (Bellman and Zadeh 1970). A further step was taken in (Zadeh 1975b). The concept of a generalized constraint was introduced in (Zadeh 1986b). A more detailed description was given in (Zadeh 1999) and in (Zadeh 2002, 2005a, 2006b).

The principal features of the concept of generalized constraint are summarized in Fig. 18 and et seq.

Constrained Variable

The constraint variable, X , can assume a variety of forms. Among the principal forms are the following.

X	is an n -ary variable, $X = (X_1, \dots, X_n)$
X	is a proposition, e. g., Leslie is tall
X	is a relation
X	is a function of another variable: $X = f(Y)$
X	is conditioned on another variable, X/Y
X	is conditioned on another generalized constraint, X isr if Y iss S
X	has a structure, e. g., $X = \text{Location (residence (Carol))}$
X	is a generalized constraint, $X: Y$ isr R
X	is a group variable. In this case, there is a group, $G: (\text{Name}_1, \dots, \text{Name}_n)$, with each member of the group, Name_i , $i = 1, \dots, n$, associated with an attribute-value, h_i , of attribute H . h_i may be vector-valued. Symbolically $G = (\text{Name}_1, \dots, \text{Name}_n)$, $G[H] = (\text{Name}_1/h_1, \dots, \text{Name}_n/h_n)$, $G[H \text{ is } A] = (\mu_A(h_1)/\text{Name}_1, \dots, \mu_A(h_n)/\text{Name}_n)$. Basically, $G[H]$ is a relation and $G[H \text{ is } A]$ is a fuzzy restriction of $G[H]$ (Zadeh 1973, 1975a, 1975b) The concept of a group variable is closely related to the concept of a fuzzy relation.

Modalities of Generalized Constraints

The indexical variable, r , defines the modality of a generalized constraint, that is, its semantics. The principal modalities are listed below.

$r :=$	equality constraint: $X = R$ is an abbreviation of $X \text{ is } R$
$r : \leq$	inequality constraint: $X \leq R$
$r : \subset$	subthood constraint: $X \subset R$
$r : \text{blank}$	possibilistic constraint; $X \text{ is } R$; R is the possibility distribution of X
$r : p$	probabilistic constraint; $X \text{ isp } R$; R is the probability distribution of X
$r : v$	veristic constraint; $X \text{ isv } R$; R is the verity (truth) distribution of X
Standard constraints	bivalent possibilistic, bivalent veristic and probabilistic
$r : rs$	random set constraint; $X \text{ isrs } R$; R is the set-valued probability distribution of X
$r : fg$	fuzzy graph constraint; $X \text{ isfg } R$; X is a function and R is its fuzzy graph
$r : u$	usuality constraint; $X \text{ isu } R$ means usually ($X \text{ is } R$)
$r : g$	group constraint; $X \text{ isg } R$ means that R constrains the attribute-values of the group
$r : i$	iterated constraint; $X \text{ is } R$ means that $X \text{ iss } S$ and $S \text{ ist } T$.

To define the semantics of various modalities it is convenient to assume that the constrained variable, X , takes values in a finite set $U = (u_1, \dots, u_n)$. With this assumption, the semantics of various constraints may be defined as follows.

Possibilistic Constraint

Consider the possibilistic constraint

$$X \text{ is } A,$$

where A is a fuzzy set in U , defined as (Zadeh 1973, 1975a).

$$A = \mu_1/u_1 + \dots + \mu_n/u_n,$$

with the understanding that $+$ is a separator and μ_i is the grade of membership of u_i in A , $i = 1, \dots, n$. The meaning of the possibilistic constraint, $X \text{ is } A$, is defined as

$$X \text{ is } A \xrightarrow{\text{definition}} \text{Poss}(X = u_i) = \mu_i, \quad i = 1, \dots, n.$$

Probabilistic Constraint

Let P be a probability distribution defined on U . P may be expressed as (Zadeh 2002, 2006b).

$$P = p_1 \setminus u_1 + \dots + p_n \setminus u_n.$$

In this case, $X \text{ isp } P \xrightarrow{\text{definition}} \text{Prob}(X = u_i) = p_i$, $i = 1, \dots, n$.

It should be noted that p_i and u_i are allowed to take granular values, P_i and U_i , respectively, meaning that

$$p_i \text{ is } P_i \text{ and } u_i \text{ is } U_i, \quad i = 1, \dots, n.$$

In this case, the probability distribution

$$P = P_1 \setminus U_1 + \dots + P_n \setminus U_n$$

is a granule-valued probability distribution in the sense defined in section “[The Concept of Granulation](#)”. Alternatively, a granule-valued distribution may be viewed as the result of granulation of the expression

$$P = p_1 \setminus u_1 + \dots + p_n \setminus u_n.$$

A granular probability distribution may be defined as an iterated generalized constraint. More specifically,

$$X \text{ isp } P$$

$$P : \text{Prob}(X \text{ is } A_i) \text{ is } P_i, \quad i = 1, \dots, n,$$

where the A_i are granules of X and the P_i are granular probabilities (Fig. 5). In this case, as in the case of granule-valued distributions, P is expressed as

$$P = P_1 \setminus U_1 + \dots + P_n \setminus U_n.$$

Note. Two examples will help to clarify the distinction between granular probability distributions and granule-valued probability distributions.

Example (a): Granular probability distribution. X is a real-valued random variable with probability distribution P . What is known about P is: $\text{Prob}(X \text{ is small})$ is low; $\text{Prob}(X \text{ is medium})$ is high; $\text{Prob}(X \text{ is large})$ is low. Example (b):

Granule-valued probability distribution. X is a random variable taking the values small, medium and large with respective granular probabilities low, high and low. Question: What are the expected values of these probability distributions?

Note. The concepts of granular and granule-valued probability distributions are closely related to the concepts of granular and granule-valued possibility distributions.

Veristic Constraint (Zadeh 1999, 2005a)

In this case, the semantics of $X \text{ isv } A$ is defined by

$$X \text{ isv } A \xrightarrow{\text{definition}} \text{Ver}(X = u_i) = \mu_i, \quad i = 1, \dots, n.$$

where $\text{Ver}(X = u_i)$ is the verity (truth) of the proposition $X = u_i$.

Fuzzy Graph Constraint

In this case, X is a function, f , from U to V . Assume that U and V are granulated, with the granules of U and V being $A_1, \dots, A_m$ and $B_1, \dots, B_n$, respectively. The fuzzy graph, R , of f is defined as the disjunction of Cartesian products of the A_i and the $B_{j(i)}$ (Zadeh 1973, 1974, 1976, 1996) (Fig. 19).

$$R = A_1 \times B_{j(1)} + \dots + A_m \times B_{j(m)}.$$

The fuzzy graph constraint is defined as the possibilistic constraint

$$f \text{ isfg } R \xrightarrow{\text{definition}} f \text{ is } (A_1 \times B_{j(1)} + \dots + A_m \times B_{j(m)}).$$

Usuality Constraint (Zadeh 1999, 2002)

The constraint $X \text{ isu } A$ is defined by

$$X \text{ isu } A \xrightarrow{\text{definition}} \text{Prob}(X \text{ is } A) \text{ is usually,}$$

where usually is a fuzzy set in $[0, 1]$ which represents a fuzzy probability.

Primary Constraints

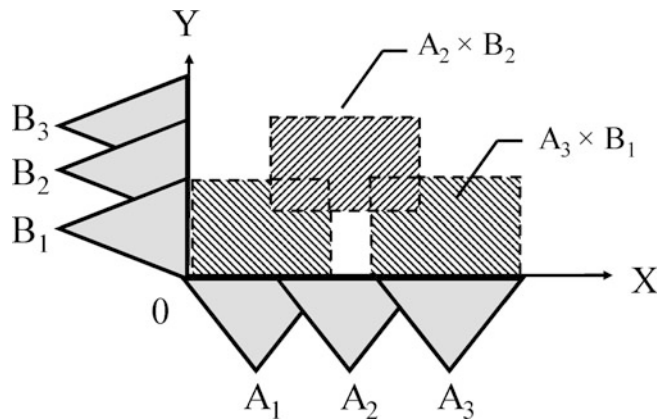
Every conceivable constraint can be viewed as an instance of a generalized constraint. In practice, such generality is rarely needed. What is sufficient for most practical purposes is a subset of generalized constraints which can be generated from so-called primary constraints through combination, projection, qualification, propagation and counterpropagation. The primary constraints are: (a) possibilistic; (b) probabilistic; and (c) veristic. The primary constraints are somewhat analogous to the primary colors: red, green and blue.

Standard Constraints

In large measure, scientific theories are based on what may be called standard constraints – constraints which are a subset of primary constraints. The standard constraints are: (a) bivalent possibilistic; (b) probabilistic; and (c) bivalent veristic.

What is important to note is that generality of generalized constraints goes far beyond the generality of standard constraints. What this points to is that the concept of a generalized constraint

Fuzzy Logic,
Fig. 19 Fuzzy graph



opens the door to a wide-ranging generalization of scientific theories.

Generalized Constraint Language (GCL)

(Zadeh 1999, 2005a)

The concept of a generalized constraint serves as a basis for construction of what is referred to as the Generalized Constraint Language (GCL). More specifically, GCL is generated by combination, projection, qualification, propagation and counter-propagation of generalized constraints. For example, combination of the probabilistic constraint

$$X \text{ is } R,$$

where X is a variable which takes values in a finite set, and the possibilistic constraint

$$(X, Y) \text{ is } S,$$

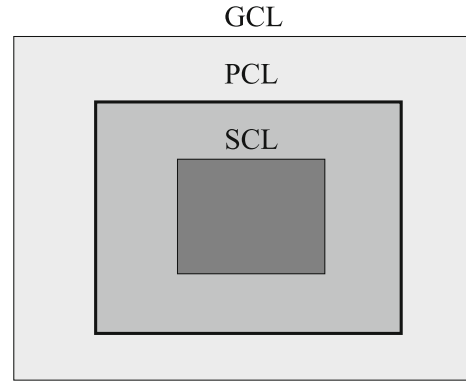
generates the fuzzy random set constraint (Zadeh 2002, 2005a, 2006b).

$$Y \text{ isfrs } T.$$

Fuzzy random sets are closely related to fuzzy-set-valued random variables. There is an extensive literature on fuzzy-set-valued random variables (Colubi et al. 2001; Goodman and Nguyen 1985; Nguyen 1993; Ogura et al. 2002; Orlov 1980; Puri and Ralescu 1993; Wang and Sanchez 1982). Random sets and set-valued random variables are closely related to the Dempster–Shafer theory of evidence (Dempster 1967; Schum 1994; Shafer 1976). An extension of the Dempster–Shafer theory to fuzzy sets and fuzzy probabilities is sketched in (Zadeh 1979a, 1981).

GCL is an open language in the sense that generalized constraints may be added to GCL at will. Simple examples of generalized constraints in GCL are:

$$\begin{aligned} &(X \text{ is } R) \text{ and } ((X, Y) \text{ is } S) \\ &(X \text{ is } R) \text{ and } (Y \text{ is } S) \\ &(X \text{ is } R) \text{ is } S \\ &(Y \text{ is } B) \text{ if } (X \text{ is } A). \end{aligned}$$



Fuzzy Logic, Fig. 20 GCL, PCL and SCL

In relation to GCL, PCL (Primary Constraint Language) and SCL (Standard Constraint Language) are subsets of GCL which are generated, respectively, by primary and standard constraints (Fig. 20).

Note. GCL is more than a language – it is a language system. A language has descriptive capability. A language system has deductive capability in addition to descriptive capability. GCL has both capabilities.

The concept of a generalized constraint plays a pivotal role in fuzzy logic. In particular, it serves to precisiate the concepts of information and meaning. More specifically, the fundamental thesis of fuzzy logic is that information may be represented as a generalized constraint.

$$\text{information} = \text{generalized constraint},$$

with the understanding that information relates to the value of a variable, X , to which the generalized constraint applies. For example, if the information about X is that X is small, then this information may be represented as a possibilistic constraint

$$X \text{ is small},$$

with small being a granular value of X . It should be noted that the traditional view that information is statistical in nature is a special, albeit important case, of viewing information as a generalized constraint. Another point which should be noted is that in information theory the primary concern is not with the substance of information but with

its measure. The fundamental thesis relates to substance rather than measure (Dubois and Prade 1994; Klir 2006).

An important corollary of the fundamental thesis is the meaning postulate of fuzzy logic. More

specifically, let p be a proposition. A proposition is a carrier of information. Consequently,

meaning of p = generalized constraint.

More concretely,

meaning of proposition = closed generalized constraint

meaning of predicate = open generalized constraint.

The meaning postulate leads to an important connection between the concept of a generalized constraint and the concept of mm-precisiation. More specifically, what can be concluded is the equality

mm – precisian = generalized constraint.

Equivalently,

mm – precisiation = translation into GCL.

This equality may be viewed as a more concrete statement of the meaning postulate. Transparency of translation may be enhanced through annotation. Details may be found in (Zadeh 2006b).

The meaning postulate points to an important aspect of translation into GCL.

Let SCL denote the subset of GCL which is associated with standard constraints, that is, bivalent possibilistic, probabilistic and bivalent veristic constraints. Let p be a proposition. An mm-precisian of p , p^* , is an element of GCL.

Let $C(p^*, p)$ be the cointension of p^* in relation to p , and let $\sup_{\text{GCL}} C(p^*, p)$ and $\sup_{\text{SCL}} C(p^*, p)$ be the suprema of $C(p^*, p)$ over GCL and SCL, respectively. Since SCL is a subset of GCL, we have

$$\sup_{\text{GCL}} C(p^*, p) \geq \sup_{\text{SCL}} C(p^*, p).$$

This obvious inequality has an important implication. Specifically, as a meaning representation language, fuzzy logic dominates bivalent logic. As a very simple example consider the

proposition p : Speed limit is 65 mph. Realistically, what is the meaning of p ? The inequality implies that employment of fuzzy logic for precisiation of p would lead to a precisian whose cointension is at least as high – and generally significantly higher – than the cointension which is achievable through the use of bivalent logic.

More concretely, assume that A tells B that the speed limit is 65 mph, with the understanding that 65 mph should be interpreted as “approximately 65 mph”. B asks A to precisiate what is meant by “approximately 65 mph”, and stipulates that no imprecise numbers and no probabilities should be used in precisiation. With this restriction, A is not capable of formulating a realistic meaning of “approximately 65 mph”. Next, B allows A to use imprecise numbers but no probabilities. B is still unable to formulate a realistic definition. Next, B allows A to employ imprecise numbers but no imprecise probabilities. Still, A is unable to formulate a realistic definition. Finally, B allows A to use imprecise numbers and imprecise probabilities. This allows A to formulate a realistic definition of “approximately 65 mph”. This simple example is intended to demonstrate the need for the vocabulary of fuzzy logic to precisiate the meaning of terms and concepts which involve imprecise probabilities.

In addition to serving as a basis for precisiation of meaning, GCL serves another important function – the function of a deductive question-answering system (Zadeh 2006a). In this role, what matters are the rules of deduction. In GCL, the rules of deduction coincide with the rules which govern constraint propagation and

counterpropagation. Basically, these are the rules which govern generation of a generalized constraint from other generalized constraints (Zadeh 2006a, 2006b).

The principal rule of deduction in fuzzy logic is the so-called extension principle (Zadeh 1965, 1975a). The extension principle can assume a variety of forms, depending on the generalized constraints to which it applies. A basic form which involves possibilistic constraints is the following. An analogous principle applies to probabilistic constraints.

Let X be a variable which takes values in U , and let f be a function from U to V . The point of departure is a possibilistic constraint on $f(X)$ expressed as

$$f(X) \text{ is } A,$$

where A is a fuzzy relation in V which is defined by its membership function $\mu_A(v)$, $v \in V$.

Let g be a function from U to W . The possibilistic constraint on $f(X)$ induces a possibilistic constraint on $g(X)$ which may be expressed as

$$g(X) \text{ is } B,$$

where B is a fuzzy relation. The question is: What is B ?

The extension principle reduces the problem of computation of B to the solution of a variational problem. Specifically,

$$\frac{f(X) \text{ is } A}{g(X) \text{ is } B}$$

where

$$\mu_B(w) = \sup_u \mu_A(f(u))$$

subject to

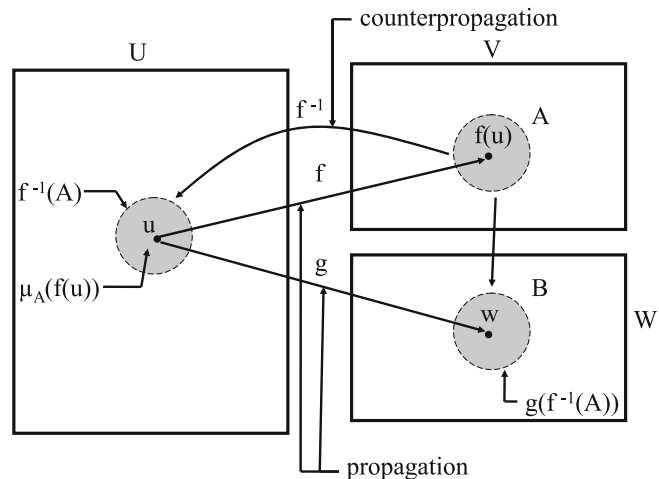
$$w = g(u).$$

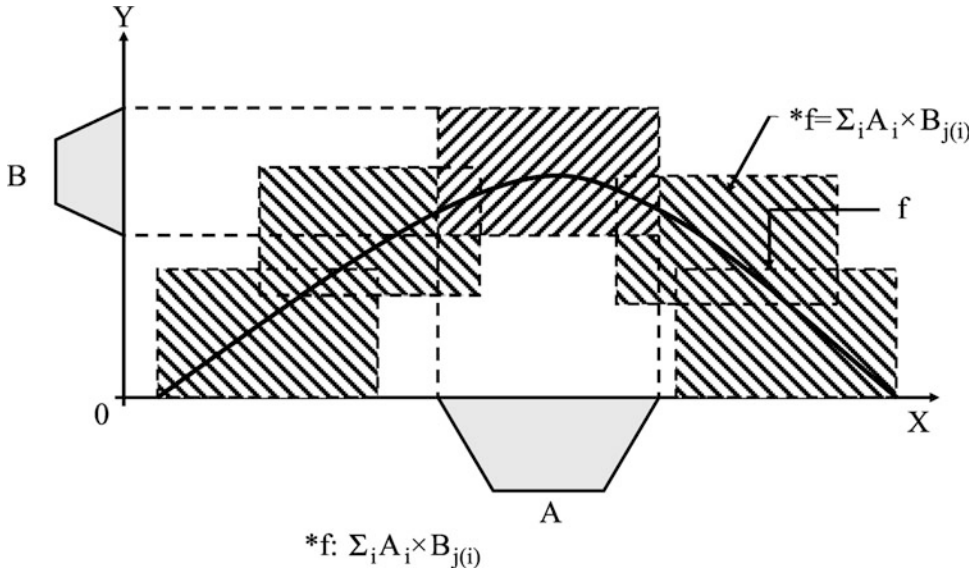
The structure of the solution is depicted in Fig. 21. Basically, the possibilistic constraint on $f(X)$ counterpropagates to a possibilistic constraint on X . Then, the possibilistic constraint on X propagates to a possibilistic constraint on $g(X)$.

There is a version of the extension principle – referred to as the fuzzy-graph extension principle – which plays an important role in control and systems analysis (Zadeh 1974, 1976, 1996). More specifically, let f be a function from reals to reals, $Y = f(X)$. Let $*f$ and $*X$ be the granulands of f and X , respectively, with $*f$ having the form of a fuzzy graph (section “The Concept of Granulation”).

Fuzzy Logic,

Fig. 21 Structure of the extension principle





Fuzzy Logic, Fig. 22 Fuzzy-graph extension principle. $B = {}^*f(A)$

$${}^*f = A_1 \times B_{j(1)} + \dots + A_m \times B_{j(m)},$$

where the A_i , $i = 1, \dots, m$ and the B_j , $j = 1, \dots, n$, are granules of X and Y , respectively; $\times$ denotes Cartesian product; and $+$ denotes disjunction (Fig. 22).

In this instance, the extension principle may be expressed as follows.

$$\frac{\begin{array}{l} X \text{ is } A \\ f \text{ is } (A_1 \times B_{j(1)} + \dots + A_m \times B_{j(m)}) \end{array}}{Y \text{ is } (m_1 \wedge B_{j(1)} + \dots + m_m \wedge B_{j(m)})}$$

where the m_i are matching coefficients, defined as (Zadeh 1976).

$$m_i = \sup(A \cap A_i), \quad i = 1, \dots, m$$

and $\wedge$ denotes conjunction (min). In the special case where X is a number, a , the possibilistic constraint on Y may be expressed as

$$Y \text{ is } (\mu_{A_1}(a) \wedge B_{j(1)} + \dots + \mu_{A_m}(a) \wedge B_{j(m)}).$$

In this form, the extension principle plays a key role in the Mamdani–Assilian fuzzy logic controller (Mamdani and Assilian 1975).

Deduction

Assume that we are given an information set, I , which consists of a system of propositions $(p_1, \dots, p_n)$. I will be referred to as the initial information set. The canonical problem of deductive question-answering is that of computing an answer to q , $\text{ans}(q|I)$, given I (Zadeh 2006a, 2006b).

The first step is to ask the question: What information is needed to answer q ? Suppose that the needed information consists of the values of the variables $X_1, \dots, X_n$. Thus,

$$\text{ans}(q|I) = g(X_1, \dots, X_n),$$

where g is a known function.

Using GCL as a meaning precisiation language, one can express the initial information set as a generalized constraint on $X_1, \dots, X_n$. In the special case of possibilistic constraints, the constraint on the X_i may be expressed as

$$f(X_1, \dots, X_n) \text{ is } A,$$

where A is a fuzzy relation.

At this point, what we have is (a) a possibilistic constraint induced by the initial information set, and (b) an answer to q expressed as

$$\text{ans}(q|I) = g(X_1, \dots, X_n),$$

with the understanding that the possibilistic constraint on f propagates to a possibilistic constraint on g . To compute the induced constraint on g what is needed is the extension principle of fuzzy logic (Zadeh 1965, 1975a).

As a simple illustration of deduction, it is convenient to use an example which was considered earlier.

Initial information set, p : most Swedes are tall

Question, q : what is the average height of Swedes?

What information is needed to compute the answer to q ? Let P be a population of n Swedes, Swede₁, ..., Swede _{n} . Let h_i be the height of Swede _{i} , $i = 1, \dots, n$. Knowing the h_i , one can express the answer to q as

$$\text{av}(h) : \text{ans}(q|p) = \frac{1}{n}(h_1 + \dots + h_n).$$

Turning to the constraint induced by p , we note that the mm-precisiand of p may be expressed as the possibilistic constraint

$$p \xrightarrow{\text{mm-precisiand}} \frac{1}{n} \sum \text{Count}(\text{tall.Swedes}) \text{ is most,}$$

where $\Sigma \text{Count}(\text{tall.Swedes})$ is the number of tall Swedes in P , with the understanding that tall Swedes is a fuzzy subset of P . Using this definition of ΣCount (Zadeh 1983b), one can write the expression for the constraint on the h_i as

$$\frac{1}{n}(\mu_{\text{tall}}(h_1) + \dots + \mu_{\text{tall}}(h_n)) \text{ is most.}$$

At this point, application of the extension principle leads to a solution which may be expressed as

$$\mu_{\text{av}(h)}(v) = \sup_h \left(\frac{1}{n} \mu_{\text{most}}(\mu_{\text{tall}}(h_1) + \dots + \mu_{\text{tall}}(h_n)) \right),$$

$$h = (h_1, \dots, h_n)$$

subject to

$$v = \frac{1}{n}(h_1 + \dots + h_n).$$

In summary, the Generalized Constraint Language is, by construction, maximally expressive. Importantly, what this implies is that, in realistic settings, fuzzy logic, viewed as a modeling language, has a significantly higher level of power and generality than modeling languages based on standard constraints or, equivalently, on bivalent logic and bivalent-logic-based probability theory.

Principal Contributions of Fuzzy Logic

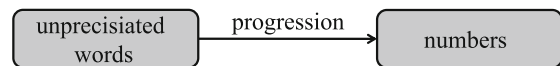
As was stated earlier, fuzzy logic is much more than an addition to existing methods for dealing with imprecision, uncertainty and complexity. In effect, fuzzy logic represents a paradigm shift. The structure of the shift is shown in Fig. 23.

Contributions of fuzzy logic range from contributions to basic sciences to applications involving various types of systems and products. The

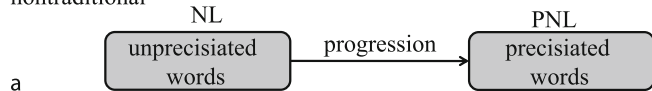
Fuzzy Logic,

Fig. 23 Fuzzy logic as a paradigm shift

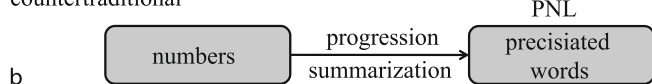
traditional



nontraditional



countertraditional



principal contributions are summarized in the following.

Fuzzy Logic as the Basis for Generalization of Scientific Theories

One of the principal contributions of fuzzy logic to basic sciences relates to what is referred to as FL-generalization.

By construction, fuzzy logic has a much more general conceptual structure than bivalent logic. A key element in the transition from bivalent logic to fuzzy logic is the generalization of the concept of a set to a fuzzy set. This generalization is the point of departure for FL-generalization.

More specifically, FL-generalization of any theory, T , involves an addition to T of concepts drawn from fuzzy logic. In the limit, as more and more concepts which are drawn from fuzzy logic are added to T , the foundation of T is shifted from bivalent logic to fuzzy logic. By construction, FL-generalization results in an upgraded theory, T^+ , which is at least as rich and, in general, significantly richer than T .

As an illustration, consider probability theory, PT – a theory which is bivalent-logic-based. Among the basic concepts drawn from fuzzy

logic which may be added to PT are the following (Zadeh 2002).

set + fuzzy set
 event + fuzzy event
 relation + fuzzy relation
 probability + fuzzy probability
 random set + fuzzy random set
 independence + fuzzy independence
 stationarity + fuzzy stationarity
 random variable + fuzzy random variable
 etc.

As a theory, PT^+ is much richer than PT. In particular, it provides a basis for construction of models which are much closer to reality than those that can be constructed through the use of PT. This applies, in particular, to computation with imprecise probabilities.

A number of scientific theories have already been FL-generalized to some degree, and many more are likely to be FL-generalized in coming years. Particularly worthy of note are the following FL-generalizations.

control $\rightarrow$ fuzzy control

(Driankov et al. 1993; Filev and Yager 1994; Gerla 2001; Ying 2000; Zadeh 1972b)

linear programming $\rightarrow$ fuzzy linear programming

(Gasimov and Yenilmez 2002; Zimmermann 1978)

probability theory $\rightarrow$ fuzzy probability theory

(Zadeh 2002, 2005, 2006)

measure theory $\rightarrow$ fuzzy measure theory

(Dubois and Prade 1982; Wang and Klir 1992)

topology $\rightarrow$ fuzzy topology

(Liu and Luo 1997; Ying 1991)

graph theory $\rightarrow$ fuzzy graph theory

[Koczy 1992; Mordeson and Nair 2000]

cluster analysis $\rightarrow$ fuzzy cluster analysis

(Bezdek et al. 1999; Hoppner et al. 1999)

Prolog $\rightarrow$ fuzzy Prolog

(Gerla 2005; Mukaidono 1989)

etc.

FL-generalization is a basis for an important rationale for the use of fuzzy logic. It is conceivable that eventually the foundations of many scientific theories may be shifted from bivalent logic to fuzzy logic.

Linguistic Variables and Fuzzy If-Then Rules The most visible, the best understood and the most widely used contribution of fuzzy logic is the concept of a linguistic variable and the associated machinery of fuzzy if-then rules (Zadeh 1996).

The machinery of linguistic variables and fuzzy if-then rules is unique to fuzzy logic. This machinery has played and is continuing to play a pivotal role in the conception and design of control systems and consumer products. However, its applicability is much broader. A key idea which underlies the machinery of linguistic variables and fuzzy if-then rules is centered on the use of information compression. In fuzzy logic, information compression is achieved through the use of graduated (fuzzy) granulation.

The Concepts of Precision and Cointension The concepts of precisiation and cointension play pivotal roles in fuzzy logic (Zadeh 2006b). In fuzzy logic, differentiation is made between two concepts of precision: precision of value, v -precision; and precision of meaning, m -precision. Furthermore, differentiation is made between precisiation of meaning which is (a) human-oriented, or mh -precisiation for short; and (b) machine-oriented, or mm -precisiation for short. It is understood that mm -precisiation is mathematically well defined. The object of precisiation, p , and the result of precisiation, p^* , are referred to as *precisiend* and *precisiand*, respectively. Informally, cointension is defined as a measure of closeness of the meanings of p and p^* . Precisiation is cointensive if the meaning of p^* is close to the meaning of p . One of the important features of fuzzy logic is its high power of cointensive precisiation. What this implies is that better models of reality can be achieved through the use of fuzzy logic.

Cointensive precisiation has an important implication for science. In large measure, science is bivalent-logic-based. In consequence, in science it is traditional to define concepts in a bivalent framework, with no degrees of truth allowed. The problem is that, in reality, many concepts in science are fuzzy, that is, are a matter of degree. For this reason, bivalent-logic-based definitions of scientific concepts are, in many cases, not cointensive. To formulate cointensive definitions of fuzzy concepts it is necessary to employ fuzzy logic.

As was noted earlier, one of the principal contributions of fuzzy logic is its high power of cointensive precisiation. The significance of this capability of fuzzy logic is underscored by the fact that it has always been, and continues to be, a basic objective of science to precisiate and clarify what is imprecise and unclear.

Computing with Words (CW), NL-Computation and Precisiated Natural Language (PNL) Much of human knowledge is expressed in natural language. Traditional theories of natural language are based on bivalent logic. The problem is that natural languages are intrinsically imprecise. Imprecision of natural languages is rooted in imprecision of perceptions. A natural language is basically a system for describing perceptions. Perceptions are intrinsically imprecise, reflecting the bounded ability of human sensory organs, and ultimately the brain, to resolve detail and store information. Imprecision of perceptions is passed on to natural languages.

Bivalent logic is intolerant of imprecision, partiality of truth and partiality of possibility. For this reason, bivalent logic is intrinsically unsuited to serve as a foundation for theories of natural language. As the logic of imprecision and approximate reasoning, fuzzy logic is a much better choice (Zadeh 1972a, 1978b, 1982, 1983a, 1983b, 1986a).

Computing with words (CW), NL-computation and precisiated natural language (PNL) are closely related formalisms (Zadeh 1999, 2000, 2001, 2004). In conventional modes of computation, the objects of computation are

mathematical constructs. By contrast, in computing with words the objects of computation are propositions and predicates drawn from a natural language. A key idea which underlies computing with words involves representing the meaning of propositions and predicates as generalized constraints. Computing with words opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories (Kacprzyk and Zadeh 1999a, 1999b; Lawry et al. 2003; Trillas 2006; Türksen 2007; Wang 2001).

Computational Theory of Perceptions Humans have a remarkable capability to perform a wide variety of physical and mental tasks without any measurements and any computations. In performing such tasks humans employ perceptions. To endow machines with this capability what is needed is a formalism in which perceptions can play the role of objects of computation. The fuzzy-logic-based computational theory of perceptions (CTP) serves this purpose (Zadeh 2000, 2001). A key idea in this theory is that of computing not with perceptions per se, but with their descriptions in a natural language. Representing perceptions as propositions drawn from a natural language opens the door to application of computing with words to computation with perceptions. Computational theory of perceptions is of direct relevance to achievement of human level machine intelligence.

Possibility Theory Possibility theory is a branch of fuzzy logic (Dubois and Prade 1988; Zadeh 1978a). Possibility theory and probability theory are distinct theories. Possibility theory may be viewed as a formalization of perception of possibility, whereas probability theory is rooted in perception of likelihood. In large measure, possibility theory and probability theory are complementary rather than competitive. Possibility theory is of direct relevance to, knowledge representation, semantics of natural languages, decision analysis and computation with imprecise probabilities.

Computation with Imprecise Probabilities Most real-world probabilities are perceptions of likelihood. As such, real-world

probabilities are intrinsically imprecise (Zadeh 1975a). Until recently, the issue of imprecise probabilities has been accorded little attention in the literature of probability theory. More recently, the problem of computation with imprecise probabilities has become an object of rapidly growing interest (Walley 1991; Zadeh 2005b).

Typically, imprecise probabilities occur in an environment of imprecisely defined variables, functions, relations, events, etc. Existing approaches to computation with imprecise probabilities do not address this reality. To address this reality what is needed is fuzzy logic and, more particularly, computing with words and the computational theory of perceptions. A step in this direction was taken in the paper “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities”; (Zadeh 2002) followed by the 2005 paper “Toward a generalized theory of uncertainty (GTU) – an outline”, (Zadeh 2005a) and the 2006 paper “Generalized theory of uncertainty (GTU) – principal concepts and ideas” (Zadeh 2006b).

Fuzzy Logic as a Modeling Language Science deals not with reality but with models of reality. More often than not, reality is fuzzy. For this reason, construction of realistic models of reality calls for the use of fuzzy logic rather than bivalent logic.

Fuzzy logic is a logic of imprecision, uncertainty and approximate reasoning (Zadeh 1979b). It is natural to employ fuzzy logic as a modeling language when the objects of modeling are not well defined (Zadeh 2008). But what is somewhat paradoxical is that in many of its practical applications fuzzy logic is used as a modeling language for systems which are precisely defined. The explanation is that, in general, precision carries a cost. In those cases in which there is a tolerance for imprecision, reduction in cost may be achieved through imprecisiation, e. g., data compression, information compression and summarization. The result of imprecisiation is an object of modeling which is not precisely defined. A fuzzy modeling language comes into play at this point. This is the

key idea which underlies the fuzzy logic gambit. The fuzzy logic gambit is widely used in the design of consumer products – a realm in which cost is an important consideration.

A Glimpse of What Lies Beyond Fuzzy Logic

Fuzzy logic has a much higher level of generality than traditional logical systems – a generality which has the effect of greatly enhancing the problem-solving capability of fuzzy logic compared with that of bivalent logic.

What lies beyond fuzzy logic? What is of relevance to this question is the so-called incompatibility principle (Zadeh 1973). Informally, this principle asserts that as the complexity of a system increases a point is reached beyond which high precision and high relevance become incompatible. The concepts of mm-precision and cointension suggest an improved version of the principle: As the complexity of a system increases a point is reached beyond which high cointension and mm-precision become incompatible. What it means in plain words is that in the realm of complex systems – such as economic systems – it may be impossible to construct models which are both realistic and precise.

As an illustration consider the following problem. Assume that *A* asks a cab driver to take him to address *B*. There are two versions of this problem. (a) *A* asks the driver to take him to *B* the shortest way; and (b) *A* asks the driver to take him to *B* the fastest way. Based on his experience, the driver decides on the route to take to *B*. In the case of (a), a GPS system can suggest a route that is provably the shortest way, that is, it can come up with a provably valid (p-valid) solution. In the case of (b) the uncertainties involved preclude the possibility of constructing a model of the system which is cointensive and mm-precise, implying that a p-valid solution does not exist. The driver's solution, based on his experience, has what may be called fuzzy validity (f-validity). Thus, in the case of (b) no p-valid solution exists. What exists is an f-valid solution.

In fuzzy logic, mm-precision is a prerequisite to computation. A question which arises is: What can be done when cointensive mm-precision is infeasible? To deal with such problems what is needed is referred to as extended fuzzy logic (FL+). In this logic, mm-precision is optional rather than mandatory.

Very briefly, what is admissible in FL+ is f-validity. Admissibility of f-validity opens the door to construction of concepts prefixed with f, e. g. f-theorem, f-proof, f-principle, f-triangle, f-continuity, f-stability, etc. An example is f-geometry. In f-geometry, figures are drawn by hand with a spray can. An example of f-theorem in f-geometry is the f-version of the theorem: The medians of a triangle are concurrent. The f-version of this theorem reads: The f-medians of an f-triangle are f-concurrent. An f-theorem can be proved in two ways. (a) empirically, that is, by drawing triangles with a spray can and verifying that the medians intersect at an f-point. Alternatively, the theorem may be f-proved by constructing an f-analogue of its traditional proof.

At this stage, the extended fuzzy logic is merely an idea, but it is an idea which has the potential for being a point of departure for construction of theories with important applications to the solution of real-world problems.

Bibliography

Primary Literature

- Aliev RA, Fazlollahi B, Aliev RR, Guirimov BG (2006) Fuzzy time series prediction method based on fuzzy recurrent neural network. In: Neuronal information processing book. Lecture notes in computer science, vol 4233. Springer, Berlin, pp 860–869
- Bargiela A, Pedrycz W (2002) Granular computing: an introduction. Kluwer Academic Publishers, Boston
- Bardossy A, Duckstein L (1995) Fuzzy rule-based modelling with application to geophysical, biological and engineering systems. CRC Press, New York
- Bellman RE, Zadeh LA (1970) Decision-making in a fuzzy environment. *Manag Sci* B 17:141–164
- Belohlavek R, Vychodil V (2006) Attribute implications in a fuzzy setting. In: Ganter B, Kwuida L (eds) ICFLA Lecture notes in artificial intelligence, vol 3874. Springer, Heidelberg, pp. 45–60

- Bezdek J, Pal S (eds) (1992) Fuzzy models for pattern recognition – methods that search for structures in data. IEEE Press, New York
- Bezdek J, Keller JM, Krishnapuram R, Pal NR (1999) Fuzzy models and algorithms for pattern recognition and image processing, Zimmermann H (ed). Kluwer, Dordrecht
- Bouchon-Meunier B, Yager RR, Zadeh LA (eds) (2000) Uncertainty in intelligent and information systems, Advances in fuzzy systems – applications and theory, vol 20. World Scientific, Singapore
- Colubi A, Santos Domínguez-Menchero J, López-Díaz M, Ralescu DA (2001) On the formalization of fuzzy random variables. *Inf Sci* 133(1–2):3–6
- Cresswell MJ (1973) Logic and languages. Methuen, London
- Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat* 38:325–329
- Driankov D, Hellendoorn H, Reinfrank M (1993) An introduction to fuzzy control. Springer, Berlin
- Dubois D, Prade H (1980) Fuzzy sets and systems – theory and applications. Academic Press, New York
- Dubois D, Prade H (1982) A class of fuzzy measures based on triangular norms. *Int J General Syst* 8:43–61
- Dubois D, Prade H (1988) Possibility theory. Plenum Press, New York
- Dubois D, Prade H (1994) Non-standard theories of uncertainty in knowledge representation and reasoning. *Knowl Engineer Rev Camb J Online* 9(4):399–416
- Esteva F, Godo L (2007) Towards the generalization of Mundici's gamma functor to IMTL algebras: the linearly ordered case, Algebraic and proof-theoretic aspects of non-classical logics, pp 127–137
- Filev D, Yager RR (1994) Essentials of fuzzy modeling and control. Wiley-Interscience, New York
- Gasimov RN, Yenilmez K (2002) Solving fuzzy linear programming problems with linear membership functions. *Turk J Math* 26:375–396
- Gerla G (2001) Fuzzy control as a fuzzy deduction system. *Fuzzy Sets Syst* 121(3):409–425
- Gerla G (2005) Fuzzy logic programming and fuzzy control. *Stud Logica* 79(2):231–254
- Godo LL, Esteva F, García P, Agustí J (1991) A formal semantical approach to fuzzy logic. In: International symposium on multiple valued logic, ISMVL'91, pp 72–79
- Goguen JA (1967) L-fuzzy sets. *J Math Anal Appl* 18: 145–157
- Goodman IR, Nguyen HT (1985) Uncertainty models for knowledge-based systems. North Holland, Amsterdam
- Hajek P (1998) Metamathematics of fuzzy logic. Kluwer, Dordrecht
- Hirota K, Sugeno M (eds) (1995) Industrial applications of fuzzy technology in the world, Advances in fuzzy systems – applications and theory, vol 2. World Scientific, Singapore
- Höppner F, Klawonn F, Kruse R, Runkler T (1999) Fuzzy cluster analysis. Wiley, Chichester
- Jamshidi M, Titli A, Zadeh LA, Boverie S (eds) (1997a) Applications of fuzzy logic – towards high machine intelligence quotient systems, Environmental and intelligent manufacturing systems series, vol 9. Prentice Hall, Upper Saddle River
- Jankowski A, Skowron A (2007) Toward rough-granular computing. In: Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, (RSFDGrC'07). Canada, Toronto, pp 1–12
- Kacprzyk J, Zadeh LA (eds) (1999a) Computing with words in information/intelligent systems part 1. Foundations. Physica, Heidelberg/New York
- Kacprzyk J, Zadeh LA (eds) (1999b) Computing with words in information/intelligent systems part 2. Applications. Physica, Heidelberg/New York
- Kandel A, Langholz G (eds) (1994) Fuzzy control systems. CRC Press, Boca Raton
- Klir GJ (2006) Uncertainty and information: foundations of generalized information theory. Wiley-Interscience, Hoboken
- Kóczy LT (1992) Fuzzy graphs in the evaluation and optimization of networks. *Fuzzy Sets Syst* 46(3): 307–319
- Lambert K, Van Fraassen BC (1970) Meaning relations, possible objects and possible worlds. In: Philosophical problems in logic. Springer, Dordrecht, pp 1–19
- Lawry J, Shanahan JG, Ralescu AL (eds) (2003) Modelling with words – learning, fusion, and reasoning within a formal linguistic representation framework. Springer, Heidelberg
- Lin TY (1997) Granular computing: From rough sets and neighborhood systems to information granulation and computing in words. In: European Congress on Intelligent Techniques and Soft Computing, September 8–12, pp 1602–1606
- Liu Y, Luo M (1997) Fuzzy topology. In: Advances in fuzzy systems – applications and theory, vol 9. World Scientific, Singapore
- Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Machine Stud* 7:1–13
- Mendel J (2001) Uncertain rule-based fuzzy logic systems – introduction and new directions. Prentice Hall, Upper Saddle River
- Mordeson JN, Nair PS (2000) Fuzzy graphs and fuzzy hypergraphs. In: Studies in fuzziness and soft computing. Springer, Heidelberg
- Mukaidono M, Shen Z, Ding L (1989) Fundamentals of fuzzy prolog. *Int J Approx Reas* 3(2):179–193
- Nguyen HT (1993) On modeling of linguistic information using random sets. In: Fuzzy sets for intelligent systems. Morgan Kaufmann Publishers, San Mateo, pp 242–246
- Novak V (2006) Which logic is the real fuzzy logic? *Fuzzy Sets Syst* 157:635–641
- Novak V, Perfilieva I, Mockor J (1999) Mathematical principles of fuzzy logic. Kluwer, Boston/Dordrecht

- Ogura Y, Li S, Kreinovich V (2002) Limit theorems and applications of set-valued and fuzzy set-valued random variables. Springer, Dordrecht
- Orlov AI (1980) Problems of optimization and fuzzy variables. Znaniye, Moscow
- Pedrycz W, Gomide F (2007) Fuzzy systems engineering: toward human-centric computing. Wiley, Hoboken
- Perfilieva I (2007) Fuzzy transforms: a challenge to conventional transforms. In: Hawkes PW (ed) Advances in images and electron physics, vol 147. Elsevier Academic Press, San Diego, pp 137–196
- Puri ML, Ralescu DA (1993) Fuzzy random variables. In: Fuzzy sets for intelligent systems. Morgan Kaufmann Publishers, San Mateo, pp 265–271
- Ralescu DA (1995) Cardinality, quantifiers and the aggregation of fuzzy criteria. Fuzzy Sets Syst 69:355–365
- Ross TJ (2004) Fuzzy logic with engineering applications, 2nd edn. Wiley, Chichester
- Rossi F, Codognet P (2003) Special Issue on Soft Constraints: Constraints 8(1)
- Rutkowska D (2002) Neuro-fuzzy architectures and hybrid learning. In: Studies in fuzziness and soft computing. Springer
- Rutkowski L (2008) Computational intelligence. Springer, Polish Scientific Publishers PWN, Warsaw
- Schum D (1994) Evidential foundations of probabilistic reasoning. Wiley, New York
- Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
- Trillas E (2006) On the use of words and fuzzy sets. Inf Sci 176(11):1463–1487
- Türksen IB (2007) Meta-linguistic axioms as a foundation for computing with words. Inf Sci 177(2):332–359
- Wang PZ, Sanchez E (1982) Treating a fuzzy subset as a projectable random set. In: Gupta MM, Sanchez E (eds) Fuzzy information and decision processes. North Holland, Amsterdam, pp 213–220
- Wang P (2001) In: Albus J, Meystel A, Zadeh LA (eds) Computing with words. Wiley, New York
- Wang Z, Klir GJ (1992) Fuzzy measure theory. Springer, New York
- Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman & Hall, London
- Wygalak M (2003) Cardinalities of fuzzy sets. In: Studies in fuzziness and soft computing. Springer, Berlin
- Yager RR, Zadeh LA (eds) (1992) An introduction to fuzzy logic applications in intelligent systems. Kluwer Academic Publishers, Norwell
- Yen J, Langari R, Zadeh LA (eds) (1995) Industrial applications of fuzzy logic and intelligent systems. IEEE, New York
- Yen J, Langari R (1998) Fuzzy logic: intelligence, control and information, 1st edn. Prentice Hall, New York
- Ying M (1991) A new approach for fuzzy topology (I). Fuzzy Sets Syst 39(3):303–321
- Ying H (2000) Fuzzy control and modeling – analytical foundations and applications. IEEE Press, New York
- Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
- Zadeh LA (1972a) A fuzzy-set-theoretic interpretation of linguistic hedges. J Cybern 2:4–34
- Zadeh LA (1972b) A rationale for fuzzy control. J Dyn Syst Meas Control G 94:3–4
- Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybern SMC 3:28–44
- Zadeh LA (1974) On the analysis of large scale systems. In: Gottinger H (ed) Systems approaches and environment problems. Vandenhoeck and Ruprecht, Göttingen, pp 23–37
- Zadeh LA (1975a) The concept of a linguistic variable and its application to approximate reasoning Part I. Inf Sci 8:199–249; Part II. Inf Sci 8:301–357; Part III. Inf Sci 9:43–80
- Zadeh LA (1975b) Calculus of fuzzy restrictions. In: Zadeh LA, Fu KS, Tanaka K, Shimura M (eds) Fuzzy sets and their applications to cognitive and decision processes. Academic Press, New York, pp 1–39
- Zadeh LA (1975c) Fuzzy logic and approximate reasoning. Synthese 30:407–428
- Zadeh LA (1976) A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. Int J Man-Machine Stud 8:249–291
- Zadeh LA (1978a) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28
- Zadeh LA (1978b) PRUF – a meaning representation language for natural languages. Int J Man-Machine Stud 10:395–460
- Zadeh LA (1979a) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) Advances in fuzzy set theory and applications. North-Holland Publishing Co., Amsterdam, pp 3–18
- Zadeh LA (1979b) A theory of approximate reasoning. In: Hayes J, Michie D, Mikulich LI (eds) Machine intelligence 9. Halstead Press, New York, pp 149–194
- Zadeh LA (1981) Possibility theory and soft data analysis. In: Cobb L, Thrall RM (eds) Mathematical frontiers of the social and policy sciences. Westview Press, Boulder, pp 69–129
- Zadeh LA (1982) Test-score semantics for natural languages and meaning representation via PRUF. In: Rieger B (ed) Empirical semantics. Brockmeyer, Bochum, pp 281–349
- Zadeh LA (1983a) Test-score semantics as a basis for a computational approach to the representation of meaning. In: Proceedings of the tenth annual conference of the association for literary and linguistic computing. Oxford University Press
- Zadeh LA (1983b) A computational approach to fuzzy quantifiers in natural languages. Comput Math 9: 149–184
- Zadeh LA (1984) Precisation of meaning via translation into PRUF. In: Vaina L, Hintikka J (eds) Cognitive constraints on communication. Reidel, Dordrecht, pp 373–402
- Zadeh LA (1986a) Test-score semantics as a basis for a computational approach to the representation of meaning. Lit Linguist Comput 1:24–35

- Zadeh LA (1986b) Outline of a computational approach to meaning and knowledge representation based on the concept of a generalized assignment statement. In: Thoma M, Wyner A (eds) *Proceedings of the international seminar on artificial intelligence and man-machine systems*. Springer, Heidelberg, pp 198–211
- Zadeh LA (1996) Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs. *Multiple-Valued Logic* 1:1–38
- Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 90:111–127
- Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Comput* 2:23–25
- Zadeh LA (1999) From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *IEEE Trans Circuits Syst* 45:105–119
- Zadeh LA (2000) Outline of a computational theory of perceptions based on computing with words. In: Sinha NK, Gupta MM, Zadeh LA (eds) *Soft computing & intelligent systems: theory and applications*. Academic Press, London, pp 3–22
- Zadeh LA (2001) A new direction in AI – toward a computational theory of perceptions. *AI Mag* 22(1):73–84
- Zadeh LA (2002) Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *J Stat Plan Inference* 105:233–264
- Zadeh LA (2004) Precisiated natural language (PNL). *AI Mag* 25(3):74–91
- Zadeh LA (2005a) Toward a generalized theory of uncertainty (GTU) – an outline. *Inf Sci* 172:1–40
- Zadeh LA (2005b) From imprecise to granular probabilities. *Fuzzy Sets Syst* 154:370–374
- Zadeh LA (2006a) From search engines to question answering systems – the problems of world knowledge, relevance, deduction and precisiation. In: Sanchez E (ed) *Fuzzy logic and the semantic web*, Chapt 9. Elsevier, pp 163–210
- Zadeh LA (2006b) Generalized theory of uncertainty (GTU)–principal concepts and ideas. *Comput Stat Data Anal* 51:15–46
- Zadeh LA (2008) Is there a need for fuzzy logic? *Inf Sci* 178(13):2751–2779
- Zimmermann HJ (1978) Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets Syst* 1:45–55
- ## Books and Reviews
- Aliev RA, Fazlollahi B, Aliev RR (2004) *Soft computing and its applications in business and economics*. In: *Studies in fuzziness and soft computing*. Springer, Berlin
- Dubois D, Prade H (eds) (1996) *Fuzzy information engineering: a guided tour of applications*. Wiley, New York
- Gupta MM, Sanchez E (1982) *Fuzzy information and decision processes*. North-Holland, Amsterdam
- Hanss M (2005) *Applied fuzzy arithmetic: an introduction with engineering applications*. Springer, Berlin
- Hirota K, Czogala E (1986) *Probabilistic sets: fuzzy and stochastic approach to decision, control and recognition processes*. ISR. Verlag TUV Rheinland, Köln
- Jamshidi M, Titli A, Zadeh LA, Boverie S (1997b) *Applications of fuzzy logic: towards high machine intelligence quotient systems*. In: *Environmental and intelligent manufacturing systems series*. Prentice Hall, Upper Saddle River
- Kacprzyk J, Fedrizzi M (1992) *Fuzzy regression analysis*. In: *Studies in fuzziness*. Physica 29
- Kosko B (1997) *Fuzzy engineering*. Prentice Hall, Upper Saddle River
- Masterakis NE (1999) *Computational intelligence and applications*. World Scientific Engineering Society
- Pal SK, Polkowski L, Skowron A (2004) *A rough-neural computing: techniques for computing with words*. Springer, Berlin
- Ralescu AL (1994) *Applied research in fuzzy technology, international series in intelligent technologies*. Kluwer Academic Publishers, Boston
- Reghis M, Roventa E (1998) *Classical and fuzzy concepts in mathematical logic and applications*. CRC-Press, Boca Raton
- Schneider M, Kandel A, Langholz G, Chew G (1996) *Fuzzy expert system tools*. Wiley, New York
- Türksen IB (2005) *Ontological and epistemological perspective of fuzzy set theory*. Elsevier Science and Technology Books
- Zadeh LA, Kacprzyk J (1992) *Fuzzy logic for the management of uncertainty*. Wiley
- Zhong N, Skowron A, Ohsuga S (1999) *New directions in rough sets, data mining, and granular-soft computing*. In: *Lecture notes in artificial intelligence*. Springer, New York

Fuzzy Probability Theory

Michael Beer
Institute for Risk and Reliability, Leibniz
University Hannover, Hannover, Germany

Article Outline

Glossary
Definition
Introduction
Mathematical Environment
Fuzzy Random Variables
Fuzzy Probability
Representation of Fuzzy Random Variables
Application Fields
Future Directions
Bibliography

Glossary

Fuzzy Set and Fuzzy Vector

Let $\underline{X}$ represent a fundamental set and $\underline{x}$ be the elements of $\underline{X}$, then

$$\tilde{A} = \{(\underline{x}, \mu_A(\underline{x})) | \underline{x} \in \underline{X}, \mu_A(\underline{x}) \geq 0 \forall \underline{x} \in \underline{X} \} \quad (1)$$

is referred to as fuzzy set $\tilde{A}$ on $\underline{X}$. $\mu_A(\underline{x})$ is the membership function (characteristic function) of the fuzzy set $\tilde{A}$ and represents the degree with which the elements $\underline{x}$ belong to $\tilde{A}$. If

$$\sup_{\underline{x} \in \underline{X}} [\mu_A(\underline{x})] = 1, \quad (2)$$

the membership function and the fuzzy set $\tilde{A}$ are called normalized; see Fig. 1. In case of a limitation to the Euclidean space $\underline{X} = \mathbb{R}^n$ and normalized fuzzy sets, the fuzzy set $\tilde{A}$ is also referred to as fuzzy vector denoted by $\tilde{x}$ with its membership

function $\mu(\underline{x})$, or, in the one-dimensional case, as fuzzy variable $\tilde{x}$ with $\mu(\underline{x})$.

α -Level Set and Support

The crisp sets

$$\underline{A}_{\alpha_k} = \{\underline{x} \in \underline{X} | \mu_A(\underline{x}) \geq \alpha_k\} \quad (3)$$

extracted from the fuzzy set $\tilde{A}$ for real numbers $\alpha_k \in (0, 1]$ are called α -level sets. These comply with the inclusion property

$$\underline{A}_{\alpha_k} \subseteq \underline{A}_{\alpha_i} \forall \alpha_i, \alpha_k \in (0, 1] \text{ with } \alpha_i \leq \alpha_k. \quad (4)$$

The largest α -level set $\underline{A}_{\alpha_k \rightarrow +0}$ is called support $S(\tilde{A})$; see Fig. 1.

σ -Algebra

A family $\mathfrak{M}(\underline{X})$ of sets $\underline{A}_i$ on the fundamental set $\underline{X}$ is referred to as σ -algebra $\mathfrak{S}(\underline{X})$ on $\underline{X}$, if

$$\underline{X} \in \mathfrak{S}(\underline{X}), \quad (5)$$

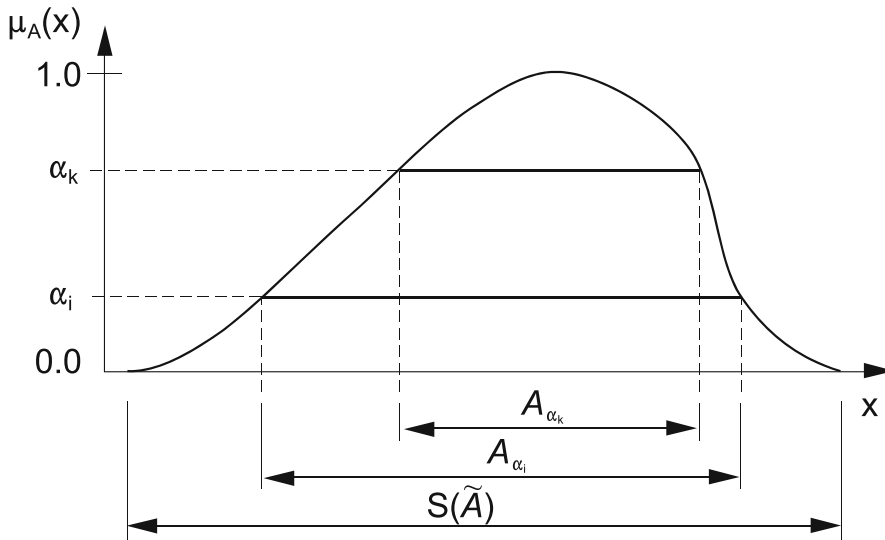
$$\underline{A}_i \in \mathfrak{S}(\underline{X}) \Rightarrow \underline{A}_i^C \in \mathfrak{S}(\underline{X}), \quad (6)$$

and if for every sequence of sets $\underline{A}_i$

$$\underline{A}_i \in \mathfrak{S}(\underline{X}); i = 1, 2, \dots \Rightarrow \bigcup_i \underline{A}_i \in \mathfrak{S}(\underline{X}). \quad (7)$$

In this definition, $\underline{A}_i^C$ is the complementary set of $\underline{A}_i$ with respect to $\underline{X}$, a family $\mathfrak{M}(\underline{X})$ of sets $\underline{A}_i$ refers to subsets and systems of subsets of the power set $\mathfrak{P}(\underline{X})$ on $\underline{X}$, and the power set $\mathfrak{P}(\underline{X})$ is the set of all subsets $\underline{A}_i$ of $\underline{X}$.

The Borel- σ -algebra $\mathfrak{B}$ is the smallest σ -algebra generated by the open sets (x_l, x_r) with $x_l, x_r \in \mathbb{Q}$, and with (x_l, x_r) forming open domains $\underline{A}_i$ in the multidimensional case.



Fuzzy Probability Theory, Fig. 1 Normalized fuzzy set with α -level sets and support

Definition

Fuzzy probability theory is an extension to probability theory to dealing with a mix of probabilistic uncertainty and non-probabilistic imprecision. It provides a theoretical basis to model uncertainty which is only partly characterized by randomness and defies a pure probabilistic modeling with certainty due to a lack of trustworthiness or precision of the data or a lack of pertinent information. The fuzzy probabilistic model is settled between probabilistic and non-probabilistic (set-theoretical) models; it covers the entire range between these models including the special bounding cases of pure randomness and pure fuzziness. The significance of fuzzy probability theory lies in the consideration of an entire set of probabilistic models, which all comply with the perception of the reality. This allows to specify the fuzzy probabilistic model as precisely as the perception allows and to leave remaining indeterminacy in the model instead of removing it artificially by way of assumptions. A fuzzy probabilistic model can be specified at the distribution level, when information about parameters, probabilities, conditions, etc. is set-valued, or it can be established in a statistical manner based on imprecise data. In the latter case, the elements of a population are not crisp but set-valued and can be understood as

granules. These cases largely comply with reality in most everyday situations. Probabilistic uncertainty and non-probabilistic imprecision can so be transferred adequately and separately to the results of a subsequent analysis. This enables best case and worst case estimates in terms of probability as well as extended sensitivity analyses taking account of the inherent non-probabilistic imprecision. The development of fuzzy probability theory was initiated by H. Kwakernaak with the introduction of fuzzy random variables (Kwakernaak 1978) in 1978. Subsequent developments have been reported in different directions and from different perspectives including differences in terminology. The usefulness of the theory has been underlined with various applications beyond mathematics and information theory, in particular, in engineering. The application fields are not limited and possess potential for extension, for example, to medicine, biology, psychology, economy, financial sciences, social sciences, and even to law.

Introduction

The probably most popular example of fuzzy probability is the evaluation of a survey on the subjective assessment of temperature. A group of

test persons are asked – under equal conditions – to give a statement on the current temperature as realistically as possible. If the scatter of the statements is considered as random, the mean value of the statements provides a reasonable statistical point estimate for the actual temperature. The statements are, however, naturally given in an imprecise form. The test persons enunciate their perception in a form such as *about 25 °C*, *possibly 27 °C*, *between 24 °C and 26 °C*, or they even only come up with linguistic assessments such as *warm*, *very warm*, or *pleasant*. This imprecision is non-probabilistic but has to be taken into account in the statistical evaluation. It is transferred to the estimated mean value, which is no longer obtained as a crisp number but as a value range or a set of values corresponding to the possibilities within the range of imprecision of the statements. If the imprecise statements are initially quantified as fuzzy values, the mean value is obtained as a fuzzy value, too; and the probability of certain events is also computed as a fuzzy value – referred to as fuzzy probability. This example is a typical materialization of the following general real-world situation.

The numerical representation of a physical variable with the aid of crisp numbers $x \in \mathbb{R}$ or sets thereof is frequently interfered by doubts regarding the trustworthiness of measured, or otherwise specified, values. The perceptions of physical variables may appear, for example, as imprecise, diffuse, vague, dubious, or ambiguous. Underlying reasons for this phenomenon include the limited precision of any measurement (digital or analog), indirect measurements via auxiliary variables in conjunction with a – more or less trustworthy – model to eventually determine the value wanted, measurements under weakly specified or arbitrarily changing boundary conditions, and the specification of values by experts in a linguistic manner. The type of the associated imprecision is non-frequentative and improper for a subjective probabilistic modeling; hence, it is non-probabilistic. This imprecision is unavoidable and may always be made evident by a respective choice of scale.

If a set of imprecise perceptions of a physical variable is present in the form of a random sample,

then the overall uncertainty possesses a mixed probabilistic/non-probabilistic character. While the scatter of the realizations of the physical variable possesses a probabilistic character (frequentative or subjective), each particular realization from the population may, additionally, exhibit non-probabilistic imprecision. Consequently, a realistic modeling in those cases must involve both probabilistic uncertainty and non-probabilistic imprecision. This modeling without distorting or ignoring information is the mission of fuzzy probability theory. A pure probabilistic modeling would introduce unwarranted information in the form of a distribution function that cannot be justified and would thus diminish the trustworthiness of the probabilistic results.

Mathematical Environment

Fuzzy probability is part of the framework of generalized information theory (Klir 2006) and is covered by the umbrella of granular computing (Lin et al. 2002; Pedrycz et al. 2008). It represents a special case of imprecise probabilities (de Cooman 2002; Walley 1991) with ties to concepts of random sets (Matheron 1975). This is underlined by Walley's summary of the semantics of imprecise probabilities with the term *indeterminacy*, which arises from ignorance about facts, events, or dependencies. Within the class of mathematical models covered by the term *imprecise probabilities* (de Cooman 2002; Klir 2006; Walley 1991), fuzzy probability theory has a relationship to concepts known as upper and lower probabilities (Hall and Lawry 2004), sets of probability measures (Fetz and Oberguggenberger 2004), distribution envelopes (Berleant and Zhang 2004), interval probabilities (Weichselberger 2000), and probability boxes (Ferson and Hajagos 2004). Also, similarities exist with respect to evidence theory (or Dempster-Shafer theory) (Dubois and Prade 1986; Shafer 1976) as a theory of infinitely monotone Choquet capacities (Klir and Folger 1988; Oberkampf et al. 2001). The relationship to the latter is associated with the interpretation of the measures plausibility and belief, with the special cases of possibility and necessity, as upper and

lower probabilities, respectively (Bernardini and Tonon 2004).

Fuzzy probability shares the common feature of all imprecise probability models: the uncertainty of an event is characterized with a set of possible measure values in terms of probability, or with bounds on probability. Its distinctive feature is that set-valued information, and hence the probability of associated events, is described with the aid of imprecise sets according to fuzzy set theory (Zadeh 1965; Zimmermann 1992). This represents a marriage between fuzzy set theory and probability theory with fuzziness and randomness as special cases, which justifies the denotation as *fuzzy randomness*. Fuzzy probability theory enables a consideration of a fuzzy set of possible probabilistic models over the range of imprecision of the knowledge about the underlying randomness. The associated fuzzy probabilities provide weighted bounds on probability – the weights of which are obtained as the membership values of the fuzzy sets. Based on α -discretization (Zimmermann 1992) and the representation of fuzzy sets as sets of α -level sets, the relationship of fuzzy probability theory to various concepts of imprecise probabilities becomes obvious. For each α -level, a common interval probability, crisp bounds on probability, or a classical set of probability measures, respectively, are obtained. Hence, for a selected α -level, the fuzzy probabilistic model coincides with a probability box. The α -level sets of fuzzy events can be treated as random sets. Further, a relationship of these random sets to evidence theory can be constructed if a respective basic probability assignment is selected; see Fellin et al. (2005). Consistency with evidence theory is obtained if the focal sets are specified as fuzzy elementary events and if the basic probability assignment follows a discrete uniform distribution over the fuzzy elementary events. These relationships are described in detail in Beer et al. (2013a) in an engineering context. The model fuzzy randomness with its two components – fuzziness and randomness – can utilize a distinction between aleatory uncertainty and non-probabilistic epistemic uncertainty with respect to the sources of uncertainty (Eldred et al. 2011; Helton et al. 2004). This can be helpful

in view of practical applications. Irreducible uncertainty as a property of the system associated with fluctuations/variability may be summarized as aleatory uncertainty and described probabilistically, and reducible uncertainty as a property of the analysts, or of the perception, associated with a lack of knowledge/precision may be understood as epistemic uncertainty and described with fuzzy sets. Detailed explanations in the wider context of imprecise probabilities are provided in (Beer et al. 2013a). The model fuzzy randomness then combines, without mixing, both components in the form of a fuzzy set of possible probabilistic models over some particular range of imprecision. This distinction is retained throughout any subsequent analysis and reflected in the results.

The development of fuzzy probability was initiated with the introduction of fuzzy random variables by Kwakernaak (1978, 1979) in 1978/1979. Subsequent milestones were set by Puri and Ralescu (1986), Kruse and Meyer (1987), Wang and Zhang (1992), and Krätschmer (2001). The developments show differences in terminology, concepts, and in the associated consideration of measurability; and the investigations are ongoing (Colubi et al. 2001, 2002a; Couso and Sanchez 2008; Diamond and Kloeden 1994; Kim 2002; Klement et al. 1986; Klement 1991; Krätschmer 2004; Terán 2006). Krätschmer (2001) showed that the different concepts can be unified to a certain extent. Generally, it can be noted that α -discretization is utilized as a helpful instrument. An overview with specific comments on the different developments is provided in (Krätschmer 2001; Gil et al. 2006). Investigations were pursued on independent and dependent fuzzy random variables for which parameters were defined with particular focus on variance and covariance (Couso and Dubois 2009; Feng et al. 2001; Hwang and Yao 1996; Körner 1997a, b; Näther and Wünsche 2007). Fuzzy random processes were examined to reveal properties of limit theorems and martingales associated with fuzzy randomness (Krätschmer 2002; Puri and Ralescu 1991; see also Song et al. 1997; Wang and Zhong 1994) and for a survey (Li et al. 2002). Particular interest was devoted to the strong law of large numbers (Colubi et al. 1999; Jang and Kwon

1998; Joo and Kim 2001). Further, the differentiation and the integration of fuzzy random variables was investigated in (López-Díaz and Gil 1998; Puri and Ralescu 1983).

Driven by the motivation for the establishment of fuzzy probability theory, considerable effort was made in the modeling and statistical evaluation of imprecise data. Fundamental achievements were reported by Kruse and Meyer (1987), Bandemer and Näther (1992; Bandemer 1992), and by Viertl (1996, 2011). Classical statistical methods were extended in order to take account of statistical fluctuations/variability and imprecision simultaneously, and the specific features associated with the imprecision of the data were investigated. Further research is reported, for example, in (Beer 2009; Näther and Körner 2002; Terán 2007) in view of evaluating measurements, in (López-Díaz and Gil 1998; Pedroni et al. 2013; Rodríguez-Muñiz et al. 2005) for decision-making, and in (Diamond 1990; Gonzalez-Rodriguez et al. 2009; Körner and Näther 1998; Näther 2006; Schollmeyer and Augustin 2015) for regression analysis. In view of applications, for example to characterize system response characteristics, the approximation of functions based on imprecise data has been investigated. In (Utkin 2019), an imprecise support vector machine is proposed for decision-making, and a neural network trained on imprecise data is developed in (Sadeghi et al. 2019). Methods for evaluating imprecise data with the aid of generalized histograms are discussed in (Bodjanova 2000; Viertl and Trutschnig 2006). In (Crespo et al. 2019), a sliced normal class of distributions is proposed to represent multivariate data, which can capture complex parameter dependencies effectively. Also, the application of resampling methods is pursued; bootstrap concepts are utilized for statistical estimations (Ferraro et al. 2011; Hung 2001) and hypothesis testing (González-Rodríguez et al. 2006; Ramos-Guajardo et al. 2010) based on imprecise data. Another method for hypothesis testing is proposed in (Grzegorzewski 2000), which employs fuzzy parameters in order to describe a fuzzy transition between rejection and acceptance. Bayesian methods have also been extended by the inclusion of fuzzy variables to take account of imprecise data;

see Viertl (1996, 2011) for comprehensive explanations. A contribution to Bayesian statistics with imprecise prior distributions is presented in Viertl and Hareter (2004). This leads to imprecise posterior distributions, imprecise predictive distributions, and may be used to deduce imprecise confidence intervals. An application of a nonparametric Bayesian approach with sets of priors is demonstrated in (Walter et al. 2017) for estimating the reliability of a system. The influence of imprecision in the priors on the results of a Bayesian update is examined quantitatively in (Zhang and Shields 2018a). Various forms of fuzzification of the Bayesian update, combinations thereof and a numerical investigation on the propagation of fuzziness through the Bayesian update are discussed in (Stein et al. 2013). A combination of the Bayesian theorem with kriging based on imprecise data is described in (Bandemer and Gebhardt 2000). A Bayesian test of fuzzy hypotheses is discussed in (Taheri and Behboodian 2001), while in (Samarasooriya and Varshney 2000) the application of a fuzzy Bayesian method for decision-making is presented.

In view of practical applications, probability distribution functions are defined for fuzzy random variables (Möller and Beer 2004; Viertl and Trutschnig 2006; Viertl 2011; Zhong et al. 1994). These distribution functions can easily be formulated and used for further calculations, but they do not uniquely describe a fuzzy random variable (Beer 2007). This theoretical problem is, however, generally without an effect in practical applications so that stochastic simulations may be performed according to the distribution functions (Zhang et al. 2010, 2013). Alternative simulation methods were proposed based on parametric (Colubi et al. 2002b) and nonparametric (Beer 2007; Möller and Reuter 2007) descriptions of imprecision. The approach according to (Möller and Reuter 2007) enables a direct generation of fuzzy realizations based on an incremental representation of fuzzy random variables. In (González-Rodríguez et al. 2009), two simulation approaches using support functions are presented and discussed in view of practical application. Further, simulation methods from related concepts of imprecise probabilities, such as (Auer et al. 2010; Limbourg and de Rocquigny 2010;

Limbourg et al. 2010; Trutschnig 2008), can be adopted for the simulation of fuzzy random variables using an α -level approach.

This variety of theoretical developments provides reasonable margins for the formulation of fuzzy probability theory but does not allow the definition of a unique concept. Choices have to be made within the elaborated margins depending on the envisaged application and environment. For the subsequent sections, these choices are made in view of a broad spectrum of possible applications, for example, in civil/mechanical engineering (Beer et al. 2013a; Möller and Beer 2004). These choices concern the following three main issues.

First, measurability has to be ensured according to a sound concept. According to (Krätschmer 2001), the concepts of measurable bounding functions (Kwakernaak 1978, 1979), of measurable mappings of α -level sets (Puri and Ralescu 1986), and of measurable fuzzy valued mappings (Diamond and Kloeden 1994; Klement et al. 1986) are available; or the unifying concept proposed in (Krätschmer 2001) itself, which utilizes a special topology on the space of fuzzy realizations, may be selected. From a practical point of view, the concept of measurable bounding functions is most suitable due to its analogy to traditional probability theory. On this basis, a fuzzy random variable can be regarded as a fuzzy set of traditional, real-valued random variables, each one carrying a certain membership degree. Each of these real-valued random variables is then measurable in the traditional fashion, and their membership degrees can be transferred to the respective measure values. The set of the obtained measure values including their membership degrees then represents a fuzzy probability.

Second, a concept for the integration of a fuzzy-valued function has to be selected from the different available approaches (Zimmermann 1992). This is associated with the computation of the probability of a fuzzy event. An evaluation in a mean sense weighted by the membership function of the fuzzy event is suggested in

(Zadeh 1968), which leads to a scalar value for the probability. This evaluation is associated with the interpretation that an event may occur partially. An approach for calculating the probability of a fuzzy event as a fuzzy set is proposed in (Yager 1984); the resulting fuzzy probability then represents a set of measure values with associated membership degrees. This complies with the interpretation that the occurrence of an event is binary but it is not clearly indicated if the event has occurred or not. The imprecision is associated with the observation rather than with the event. The latter approach corresponds with the practical situation in many cases, provides useful information in form of the imprecision reflected in the measure values, and follows the selected concept of measurability. It is thus taken as a basis for further consideration.

Third, the meaning of the distance between fuzzy sets as realizations of a fuzzy random variable must be defined, which is of particular importance for the definition of the variance and of further parameters of fuzzy random variables. An approach that follows a set-theoretical point of view and leads to a crisp distance measure is presented in (Körner 1997a). It is proposed to apply the Hausdorff metric to the α -level sets of a fuzzy set and to average the results over the membership scale. Consequently, parameters of a fuzzy random variable which are associated with a distance between fuzzy realizations reflect the variability within the imprecision in an integrated form. For example, the variance of a fuzzy random variable is obtained as a crisp value. In contrast to this, the application of standard algorithms for operations on fuzzy sets, such as the extension principle (Bandemer and Gottwald 1995; Klir and Folger 1988; Zimmermann 1992), leads to fuzzy distances between fuzzy sets. Parameters, including variances, of fuzzy random variables are then obtained as fuzzy sets of possible parameter values. This corresponds to the interpretation of fuzzy random variables as fuzzy sets of traditional, real-valued random quantities. The latter approach is thus pursued further.

These three selections comply basically with the definitions in Kruse and Meyer (1987); Kwakernaak (1978); see also Gil et al. (2006). Among all the possible choices, this set-up ensures the most compatible settlement of fuzzy probability theory within the framework of imprecise probabilities (de Cooman 2002; Klir 2006; Walley 1991) with ties to evidence theory and random set approaches (Fellin et al. 2005; Helton and Oberkampf 2004), see also Beer et al. (2013a). Fuzzy probability is obtained in the form of weighted plausible bounds on probability. Moreover, in view of practical applications, the treatment of fuzzy random variables as fuzzy sets of real-valued random variables enables the utilization of established probabilistic methods as kernel solutions in the environment of a fuzzy analysis. For example, sophisticated methods of Monte Carlo simulation may be combined with a generally applicable fuzzy analysis based on a global optimization approach using α -discretization (Möller et al. 2000; Möller and Beer 2004). This concept is supported by the study in (Eldred et al. 2011), where three modeling approaches for imprecise probabilities were compared: interval-valued probability, second-order probability, and evidence theory. It was found that a combination of stochastic expansions for probabilistic uncertainty with an optimization approach to determine interval bounds for probabilities (which is applicable to the α -levels of a fuzzy random variable) provides advantages in terms of accuracy and efficiency. Alternatively, if some restricting conditions are complied with, numerically efficient methods from interval mathematics (Alefeld and Herzberger 1983) may be employed for the α -level mappings instead of a global optimization approach; see Degrauwe et al. (2010); Muhanna et al. (2007). The selected concept, eventually, enables best-case and worst-case studies within the range of possible probabilistic models. The increasing complexity of real-world problems has pushed the developments towards combining the concept of global optimization with ideas from interval mathematics in order to circumvent or simplify the optimization to calculate bounds through efficient and effective problem formulations. Considerations on how to obtain tight bounds on probabilistic results in relation to the

computational effort are provided in Fetz and Oberguggenberger (2016); Wang et al. (2018); Alvarez et al. (2017); Crespo et al. (2013). Concepts for reducing the sample size are reported in Ref. (Zhang et al. 2013) using low-discrepancy sequences and in (Zhang and Shields 2018b) based on multi-model inference. Combinations with subset sampling (Alvarez et al. 2018) and importance sampling (Troffaes 2018) have shown a significant gain in numerical efficiency. Series expansions (Muscolino and Sofi 2017; Feng et al. 2017; Schöbi and Sudret 2017) have shown benefits to approximate bounds on stochastic responses of systems and structures. Expressing the failure probability as a series of the parameters of probabilistic models of the input opens an efficient pathway for performance, reliability, and sensitivity analysis in a nonintrusive manner in combination with any of the most efficient stochastic sampling concepts (Wei et al. 2019a, b; Song et al. 2019). A specific development of this concept for line sampling is presented in (Song et al. 2020). Using Bernstein polynomials enables an efficient determination of probability bounds processing hyper-cuboids of input parameters with the option for iterative refinement of the bounding of the result (Roberto et al. 2018). Intervening variables have been proposed to moderate down nonlinearities so that linear series expansions deliver high quality results at low effort (Valdebenito et al. 2020). Exploiting the topology of the problem in some parameter space has emerged as a pathway to pre-identify parameter configurations that lead directly to the bounds of the result. This concept is used in particular for reliability assessment. In (Angelis et al. 2015), the pre-identification is realized in the standard normal space based on the location of the failure domain used for line sampling. In Ref. (Diego et al. 2014), the properties of the design point are exploited, and in (Hurtado 2013), the properties of the probability plot are utilized for the pre-identification. Surrogate models provide another pathway for efficient analysis with imprecise probabilities. Achievements have been reported in particular based on adaptive kriging (Yun et al. 2019), also in combination with multilevel models (Schöbi and Sudret 2017). An approach using an interval predictor is reported in (Sadeghi et al. 2020). While these developments are

focused on reliability analysis, a neural network for optimization purposes is proposed in (Freitag et al. 2020). For systems reliability, efficient problem representations based on the concept of survival signature have been proposed, which allow a decoupling of the system structure and probabilistic structure. In this regard, imprecise probabilities can be processed efficiently in the probabilistic structure while the system is analyzed independently (Patelli et al. 2017; Coolen and Coolen-Maturi 2016). This eases the analysis of complex systems, common cause failures (George-Williams et al. 2019), and couples systems (Behrendorf et al. 2019). Expansions of the imprecise probabilities to imprecise stochastic processes and fields have been reported for response characterization and reliability analysis. The time-dependent response of a nonlinear structure is analyzed in (Graf et al. 2015) with a fuzzy probabilistic process description. A time-dependent reliability analysis is reported in (Do et al. 2014). The implementation of imprecise random fields in a stochastic finite element framework is considered in (Do et al. 2016; Faes and Moens 2019; Oberguggenberger 2015; Pivovarov et al. 2019; Wei et al. 2018), where spectral approaches are typically used for the representation and processing. For cases where the physics of the problem are not known or only very vaguely known so that no rigorous mathematical description is feasible, concepts of credal networks offer an intuitive instrument to model complex problems taking into account imprecise probabilistic information. This approach is particularly useful for inference, sensitivity, and risk analysis, as well as for decision-making (Tolo et al. 2018; Antonucci et al. 2015).

Fuzzy Random Variables

With the above selections, the definitions from traditional probability theory can be adopted and extended to dealing with imprecise outcomes from a random experiment. Let Ω be the space of random elementary events $\omega \in \Omega$ and the universe on which the realizations are observed be the n -dimensional Euclidean space $\underline{\mathbf{X}} = \mathbb{R}^n$. Then, a membership scale μ is introduced

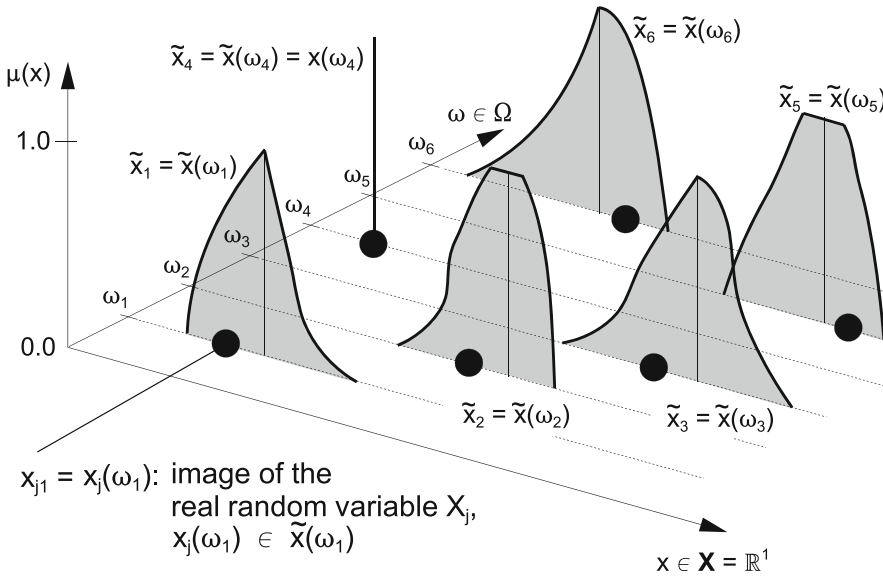
perpendicular to the hyperplane $\Omega \times \underline{\mathbf{X}}$. This enables the specification of fuzzy sets on $\underline{\mathbf{X}}$ for given elementary events ω from Ω without interaction between Ω and μ . That is, randomness induced by Ω and fuzziness described by μ – only in x -direction – are not mixed with one another. Let $\mathfrak{F}(\underline{\mathbf{X}})$ be the set of all fuzzy variables on $\underline{\mathbf{X}} = \mathbb{R}^n$; that is, $\mathfrak{F}(\underline{\mathbf{X}})$ denotes the collection of all fuzzy sets $\tilde{\underline{\mathbf{A}}}$ on $\underline{\mathbf{X}} = \mathbb{R}^n$, with $\tilde{\underline{\mathbf{A}}}$ according to Eq. 1. Then, the mapping

$$\tilde{\underline{\mathbf{X}}} : \Omega \rightarrow \mathfrak{F}(\underline{\mathbf{X}}) \quad (8)$$

is referred to as fuzzy random variable $\tilde{\underline{\mathbf{X}}}$. In contrast to real-valued random variables, a fuzzy image $\tilde{\underline{\mathbf{x}}}(\omega) \in \mathfrak{F}(\underline{\mathbf{X}})$, or $\tilde{\underline{\mathbf{x}}}(\omega) \subseteq \underline{\mathbf{X}}$, is now assigned to each elementary event $\omega \in \Omega$; see Fig. 2. These fuzzy images may be understood as a numerical representation of granules. Generally, a fuzzy random variable can be discrete or continuous with respect to both fuzziness and randomness. The further consideration refers to the continuous case, from which the discrete case may be derived.

Without a limitation in generality, the fuzzy image may be restricted to normalized fuzzy variables, thus representing fuzzy vectors. Further restrictions can be defined in view of a convenient numerical treatment if the application field allows for. This concerns, for example, a restriction to connected and compact α -level sets $\underline{\mathbf{A}}_\alpha = \underline{\mathbf{x}}_\alpha$ of the fuzzy images, a restriction to convex fuzzy sets as fuzzy images (a fuzzy set $\tilde{\underline{\mathbf{A}}}$ is convex if all its α -level sets $\underline{\mathbf{A}}_\alpha$ are convex sets), or a restriction to fuzzy images with only one element $\underline{\mathbf{x}}_i$ carrying the membership $\mu(\underline{\mathbf{x}}_i) = 1$ as in (Kwakernaak 1978).

For the treatment of the fuzzy images, a respective algorithm for operations on fuzzy sets has to be selected. Following the literature and the above selection, the standard basis is employed with the min-operator as a special case of a t-norm and the associated max-operator as a special case of a t-co-norm (Dubois and Prade 1980; Dubois and Prade 1985; Zimmermann 1992). This leads to the min-max operator and the extension principle (Bandemer and Gottwald 1995; Klir and Folger



Fuzzy Probability Theory, Fig. 2 Fuzzy random variable

1988; Zimmermann 1992). With the interpretation of a fuzzy image $\underline{x}_i$ as a fuzzy set $\tilde{x}_i = \{\underline{x}_{ji}, \mu(\underline{x}_{ji})\}$ with the above selections, a fuzzy random variable can be represented as a fuzzy set of real-valued random variables, which enables a coherent relationship to traditional probability theory. Let $\underline{x}_{ji}$ be the image of the elementary event ω_i of a real-valued random variable $\underline{X}_j$ and $\tilde{x}_i$ be the fuzzy image of a fuzzy random variable $\tilde{\underline{X}}$ with $\underline{x}_{ji}$ and $\tilde{x}_i$ be assigned to the same elementary event ω_i . If $\underline{x}_{ji} \in \tilde{x}_i$, then $\underline{x}_{ji}$ is called contained in $\tilde{x}_i$. If, for all elementary events $\omega_i \in \Omega, i = 1, 2, \dots$, the $\underline{x}_{ji}$ are contained in the $\tilde{x}_i$, the set of the $\underline{x}_{ji}, i = 1, 2, \dots$, then constitute an original $\underline{X}_j$ of the fuzzy random variable $\tilde{\underline{X}}$; see Fig. 2. The original $\underline{X}_j$ is referred to as completely contained in $\tilde{\underline{X}}$, $\underline{X}_j \in \tilde{\underline{X}}$. Each real-valued random variable $\underline{X}$ that is completely contained in $\tilde{\underline{X}}$ is an original $\underline{X}_j$ of $\tilde{\underline{X}}$ and carries the membership

$$\mu(\underline{X}_j) = \max [\alpha | \underline{x}_{ji} \in \underline{x}_{i\alpha} \forall i]. \quad (9)$$

That is, in the Ω -direction, each original $\underline{X}_j$ must be consistent with the fuzziness of $\tilde{\underline{X}}$. Consequently, the fuzzy random variable $\tilde{\underline{X}}$ can be

represented as the fuzzy set of all originals $\underline{X}_j$ contained in $\tilde{\underline{X}}$,

$$\tilde{\underline{X}} = \{(\underline{X}_j, \mu(\underline{X}_j)) | \underline{x}_{ji} \in \tilde{x}_i \forall i\}. \quad (10)$$

The representation of a fuzzy random variable as a fuzzy set of real-valued random variables according to Eq. 10 enables a workable interpretation of a fuzzy random variable in view of measuring probabilities of events on $\mathbb{R}^n$ in correspondence with traditional probability theory. This interpretation can be understood as an extension of the definition of a real-valued random variable in the following manner. Let $\underline{X}$ be a real-valued random variable defined as a measurable function

$$\underline{X} = \Omega \rightarrow \mathbb{R}^n, \quad (11)$$

which connects the probability space $[\Omega; \mathcal{F}; P]$ defined on Ω (with $\mathcal{F}$ being a σ -algebra on Ω) and the measure space $[\underline{X}; \mathfrak{B}]$ defined on $\underline{X}$, where $\mathfrak{B}$ is the Borel- σ -algebra. That is, for every subset $\underline{A}_i \in \mathfrak{B}$, the pre-image is $\underline{X}^{-1}(\underline{A}_i) \in \mathcal{F}$, for which $\underline{X}^{-1}(\underline{A}_i) = \{\omega : \underline{X}(\omega) \in \underline{A}_i\}$. Hence, the sets $\underline{A}_i$ are measured based on the measure of their pre-images being elements of $\mathcal{F}$. Then, a fuzzy random

variable $\tilde{X}$ according to Eq. 8 can be understood as a multivalued mapping from Ω to $\mathbb{R}^n$ since $\mathfrak{F}(\underline{X})$ set of (fuzzy) subsets of $\mathbb{R}^n$. This multivalued mapping facilitates multiple destinations on from the same origin on Ω , each of which carries an associated membership degree. That is, a fuzzy random variable can be understood as a fuzzy set of measurable functions that have the same origin but different destinations associated with membership degrees. Hence, for every subset $\underline{A}_i \in \mathfrak{B}$, the pre-image is $\left\{ \left(\underline{X}_j^{-1}(\underline{A}_i), \mu(\underline{X}_j^{-1}(\underline{A}_i)) \right) \right\}$ with $\underline{X}_j^{-1}(\underline{A}_i) \in \mathcal{F} \forall \underline{X}_j^{-1}(\underline{A}_i) \in \left\{ \underline{X}_j^{-1}(\underline{A}_i) \right\}$ and $\mu(\underline{X}_j^{-1}(\underline{A}_i)) = \mu(\underline{X}_j)$. Further, $\underline{X}_j^{-1}(\underline{A}_i) = \{ \omega : \underline{X}_j(\omega) \in \underline{A}_i \} \forall \underline{X}_j^{-1}(\underline{A}_i) \in \left\{ \underline{X}_j^{-1}(\underline{A}_i) \right\}$. From this consideration is obvious that traditional probability theory can be used to analyze fuzzy random variables. However, sets $\underline{A}_i \in \mathfrak{B}$ cannot be measured precisely since the pre-images $\left\{ \left(\underline{X}_j^{-1}(\underline{A}_i), \mu(\underline{X}_j^{-1}(\underline{A}_i)) \right) \right\}$ are fuzzy sets. This implies that the imprecision of the image of a fuzzy random variable is transferred to the probability of events on $\mathbb{R}^n$.

Each fuzzy random variable $\tilde{X}$ contains at least one real-valued random variable $\underline{X}$ as an original $\underline{X}_j$ of $\tilde{X}$. Each fuzzy random variable $\tilde{X}$ that possesses precisely one original is thus a real-valued random variable $\underline{X}$. That is, real-valued random variables are a special case of fuzzy random variables. This enables a simultaneous treatment of real-valued random variables and fuzzy random variables within the same theoretical environment and with the same numerical algorithms. Or, vice versa, it enables the utilization of theoretical results and established numerical algorithms from traditional probability theory within the framework of fuzzy probability theory.

If α -discretization is applied to the fuzzy random variable $\tilde{X}$, random α -level sets $\underline{X}_\alpha$ are obtained,

$$\underline{X}_\alpha = \{ \underline{X} = \underline{X}_j | \mu(\underline{X}_j) \geq \alpha \}. \quad (12)$$

Their realizations are α -level sets $\underline{x}_{i\alpha}$ of the respective fuzzy realizations $\tilde{x}_i$ of the fuzzy random variable $\tilde{X}$. A fuzzy random variable can thus, alternatively, be represented by the set of its α -level sets,

$$\tilde{X} = \{ (X_\alpha, \mu(\underline{X}_\alpha)) | \mu(\underline{X}_\alpha) = \alpha \forall \alpha \in (0, 1] \}. \quad (13)$$

In the one-dimensional case and with the restriction to connected and compact α -level sets of the realizations, the random α -level sets X_α of the fuzzy random variable $\tilde{X}$ become closed random intervals $[X_{al}, X_{ar}]$.

Fuzzy Probability

Fuzzy probability can be derived as a fuzzy set of probability measures for events the occurrence of which depends on the behavior of a fuzzy random variable. These events are referred to as fuzzy random events with the following characteristics. Let $\tilde{X}$ be a fuzzy random variable according to Eq. 8 with the realizations $\tilde{x}$ and $\mathfrak{B}$ is the Borel- σ -algebra making all sets $\underline{A}_i$ on $\underline{X}$ measurable. Then, the event

$$\tilde{E}_i : \tilde{X} \text{ hits } \underline{A}_i \quad (14)$$

is referred to as fuzzy random event, which occurs if a fuzzy realization $\tilde{x}$ of the fuzzy random variable $\tilde{X}$ hits the set $\underline{A}_i$. The associated probability of occurrence of $\tilde{E}_i$ is referred to as fuzzy probability $\tilde{P}(\underline{A}_i)$. It is obtained as the fuzzy set of the probabilities of occurrence of the events

$$E_{ij} : \underline{X}_j \in \underline{A}_i \quad (15)$$

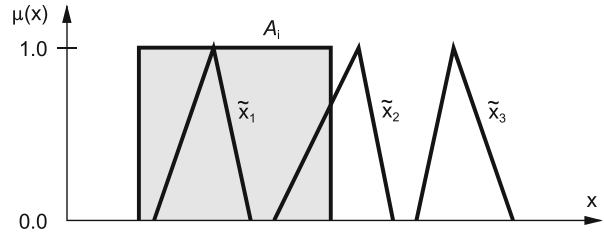
associated with all originals $\underline{X}_j$ of the fuzzy random variable $\tilde{X}$ with their membership values $\mu(\underline{X}_j)$. Specifically,

$$\begin{aligned} \tilde{P}(\underline{A}_i) = & \left\{ (P(\underline{X}_j \in \underline{A}_i), \mu(P(\underline{X}_j \in \underline{A}_i))) | \underline{X}_j \in \tilde{X}, \right. \\ & \left. \mu(P(\underline{X}_j \in \underline{A}_i)) = \mu(\underline{X}_j) \forall j \right\}. \end{aligned} \quad (16)$$

Each of the involved probabilities $P(\underline{X}_j \in \underline{A}_i)$ is a traditional, real-valued probability associated with the traditional probability space $[\Omega; \mathcal{F}; P]$ defined on Ω (with $\mathcal{F}$ being a σ -algebra on Ω) and

Fuzzy Probability

Theory, Fig. 3 Fuzzy event $\tilde{x}_k$ hits $\underline{A}_i$ in the one-dimensional case



the measure space $[X; \mathfrak{B}]$ defined on X , where $\mathfrak{B}$ is the Borel- σ -algebra, and complying with all established theorems and properties of traditional probability theory. The evaluation of the fuzzy random event Eq. 14 requires an answer to the question whether a fuzzy realization $\tilde{x}$ hits the set $\underline{A}_i$ or not. Due to the fuzziness of the $\tilde{x}$, these events appear as fuzzy events $\tilde{E}_{ik} : \tilde{x}_k \text{ hits } \underline{A}_i$ with the following three options for occurrence; see Fig. 3:

- The fuzzy realization $\tilde{x}_k$ lies completely inside the set $\underline{A}_i$, the event $\tilde{E}_{ik}$ has occurred.
- The fuzzy realization $\tilde{x}_k$ lies only partially inside $\underline{A}_i$, the event $\tilde{E}_{ik}$ may have occurred or not occurred.
- The fuzzy realization $\tilde{x}_k$ lies completely outside the set $\underline{A}_i$, the event $\tilde{E}_{ik}$ has not occurred.

The fuzzy probability $\tilde{P}(\underline{A}_i)$ takes account of all three options within its range of fuzziness. The fuzzy random variable $\tilde{X}$ is discretized into a set of random α -level sets $\underline{X}_\alpha$ according to Eq. 12, and the events $\tilde{E}_i$ and $\tilde{E}_{ik}$, respectively, are evaluated α -level by α -level. In this evaluation, the event $E_{ik\alpha} : \underline{x}_{k\alpha} \text{ hits } \underline{A}_i$ admits the following two bounding interpretations of occurrence:

- E_{ikal} : “ $\underline{x}_{k\alpha}$ is contained in $\underline{A}_i : \underline{x}_{k\alpha} \subseteq \underline{A}_i$ ”
- E_{ikar} : “ $\underline{x}_{k\alpha}$ and $\underline{A}_i$ possess at least one element in common: $\underline{x}_{k\alpha} \cap \underline{A}_i \neq \emptyset$ ”

Consequently, the events $E_{ik\alpha l}$ are associated with the smallest probability

$$P_{\alpha l}(\underline{A}_i) = P(\underline{X}_\alpha \subseteq \underline{A}_i), \quad (17)$$

and the events E_{ikar} correspond to the largest probability

$$P_{\alpha r}(\underline{A}_i) = P(\underline{X}_\alpha \cap \underline{A}_i \neq \emptyset). \quad (18)$$

The probabilities $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ are obtained as bounds on the probability $\tilde{P}(\underline{A}_i)$ on the respective α -level associated with the random α -level set $\underline{X}_\alpha$ of the fuzzy random variable $\tilde{X}$; see Fig. 4. As all elements of $\underline{X}_\alpha$ are originals $\underline{X}_j$ of $\tilde{X}$, the probability that an $\underline{X}_j \in \underline{X}_\alpha$ hits $\underline{A}_i$ is bounded according to

$$P_{\alpha l}(\underline{A}_i) \leq P(\underline{X}_j \in \underline{A}_i) \leq P_{\alpha r}(\underline{A}_i) \forall \underline{X}_j \in \underline{X}_\alpha. \quad (19)$$

This enables a computation of the probability bounds directly from the real-valued probabilities $(P(\underline{X}_j \in \underline{A}_i))$ associated with the originals $\underline{X}_j$,

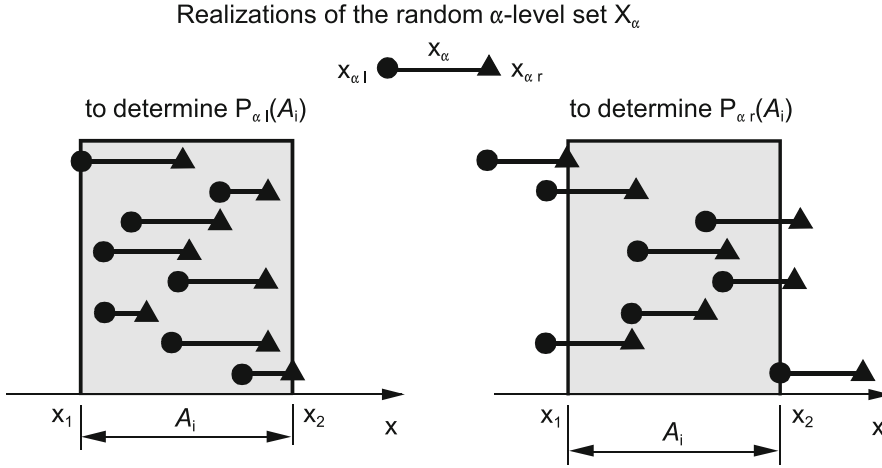
$$P_{\alpha l}(\underline{A}_i) = \min_{\underline{X}_j \in \underline{X}_\alpha} P(\underline{X}_j \in \underline{A}_i), \quad (20)$$

$$P_{\alpha r}(\underline{A}_i) = \max_{\underline{X}_j \in \underline{X}_\alpha} P(\underline{X}_j \in \underline{A}_i). \quad (21)$$

If the fuzzy random variable $\tilde{X}$ represents a fuzzy set of continuous real-valued random variables and if the membership functions of all fuzzy realizations $\tilde{x}_i$ of $\tilde{X}$ are at least segmentally continuous, then the probability bounds $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ determine closed connected intervals $[P_{\alpha l}(\underline{A}_i), P_{\alpha r}(\underline{A}_i)]$. In this case, the fuzzy probability $\tilde{P}(\underline{A}_i)$ is obtained as a continuous and convex fuzzy set, which may be specified uniquely with the aid of α -discretization,

$$\begin{aligned} \tilde{P}(\underline{A}_i) &= \{(P_\alpha(\underline{A}_i), \mu(P_\alpha(\underline{A}_i))) \mid \\ P_\alpha(\underline{A}_i) &= [P_{\alpha l}(\underline{A}_i), P_{\alpha r}(\underline{A}_i)], \\ \mu(P_\alpha(\underline{A}_i)) &= \alpha \forall \alpha \in (0, 1)\}. \end{aligned} \quad (22)$$

The properties of the fuzzy probability $\tilde{P}(\underline{A}_i)$ result from the properties of the traditional probability measure in conjunction with fuzzy set



Fuzzy Probability Theory, Fig. 4 Events for determining $P_{\alpha l}(\underline{A}_i)$ and $P_{\alpha r}(\underline{A}_i)$ in the one-dimensional case

theory. For example, a complementary relationship may be derived for $\tilde{P}(\underline{A}_i)$ as follows. The equivalence

$$(\underline{X}_x \subseteq \underline{A}_i) \Leftrightarrow (\underline{X}_x \cap \underline{A}_i^C = \emptyset) \quad (23)$$

with $\underline{A}_i^C$ being the complementary set of $\underline{A}_i$ with respect to the universe $\underline{X}$, leads to

$$P(\underline{X}_x \subseteq \underline{A}_i) = P(\underline{X}_x \cap \underline{A}_i^C = \emptyset) \quad (24)$$

for each α -level. If the event $\underline{X}_x \cap \underline{A}_i^C = \emptyset$ is expressed in terms of its complementary event $\underline{X}_x \cap \underline{A}_i^C \neq \emptyset$, Eq. 24 can be rewritten as

$$P(\underline{X}_x \subseteq \underline{A}_i) = 1 - P(\underline{X}_x \cap \underline{A}_i^C \neq \emptyset). \quad (25)$$

This leads to the relationships

$$P_{\alpha l}(\underline{A}_i) = 1 - P_{\alpha r}(\underline{A}_i^C), \quad (26)$$

$$P_{\alpha r}(\underline{A}_i) = 1 - P_{\alpha l}(\underline{A}_i^C), \quad (27)$$

and

$$\tilde{P}(\underline{A}_i) = 1 - \tilde{P}(\underline{A}_i^C). \quad (28)$$

In the special case that the set $\underline{A}_i$ contains only one element $\underline{A}_i = \underline{x}_i$, the fuzzy probability and

$\tilde{P}(\underline{A}_i)$ changes to $\tilde{P}(\underline{x}_i)$. The event $X_\alpha \cap \underline{A}_i \neq \emptyset$ is then replaced by $\underline{x}_i \in \underline{X}_x$, and $\underline{X}_x \subseteq \underline{A}_i$ becomes $\underline{X}_x = \underline{x}_i$. The probability $P_{\alpha l}(\underline{x}_i) = P(\underline{X}_x = \underline{x}_i)$ may take values greater than zero only if a realization of $\underline{X}_x$ exists that possesses exactly one element $\underline{X}_x = \underline{t}$ with $\underline{t} = \underline{x}_i$ and if this element $\underline{t}$ represents a realization of a discrete original of $\underline{X}_x$. Otherwise, $P_{\alpha l}(\underline{x}_i) = 0$, and the fuzziness of $\tilde{P}(\underline{x}_i)$ is exclusively specified by $P_{\alpha r}(\underline{x}_i)$.

In the one-dimensional case with

$$\underline{A}_i = \{x | x \in \mathbf{X}; x_1 \leq x \leq x_2\} \quad (29)$$

the fuzzy probability $\tilde{P}(\underline{A}_i)$ can be represented in a simplified manner. If the random α -level sets X_α are restricted to be closed random intervals $[X_{\alpha l}, X_{\alpha r}]$, the associated fuzzy random variable $\tilde{X}$ can be completely described by means of the bounding real-valued random variables $X_{\alpha l}$ and $X_{\alpha r}$; see Fig. 4. $X_{\alpha l}$ and $X_{\alpha r}$ represent specific originals X_j of $\tilde{X}$. This enables the specification of the probability bounds for each α -level according to

$$P_{\alpha l}(\underline{A}_i) = \max [0, P(X_{\alpha r} = t_r | x_2, t_r \in \mathbf{X}, t_r \leq x_2) - P(X_{\alpha l} = t_l | x_1, t_l \in \mathbf{X}, t_l < x_1)], \quad (30)$$

$$P_{\alpha r}(A_i) = P(X_{\alpha l} = t_l | x_2, t_l \in \mathbf{X}; t_l \leq x_2) - P(X_{\alpha r} = t_r | x_1, t_r \in \mathbf{X}; t_r < x_1). \quad (31)$$

From the above framework, the special case of a real-valued random variable $\underline{X}$ appears as a fuzzy random variable $\tilde{\underline{X}}$ that contains precisely one original $\underline{X}_j = \underline{X}_1$,

$$\underline{X}_1 = (\underline{X}_{j=1}, \mu(\underline{X}_{j=1}) = 1); \quad (32)$$

see Eq. 10. Then, all $\underline{X}_x$ contain only the sole original $\underline{X}_1$, and both $\underline{X}_x \cap \underline{A}_i \neq \emptyset$ and $\underline{X}_x \subseteq \underline{A}_i$ reduce to $\underline{X}_1 \in \underline{A}_i$. That is, the event $\underline{X}_1$ hits $\underline{A}_i$ does no longer provide options for interpretation reflected as fuzziness,

$$P_{\alpha l}(\underline{A}_i) = P_{\alpha r}(\underline{A}_i) = P(\underline{A}_i) = P(\underline{X}_1 \in \underline{A}_i). \quad (33)$$

In the above special one-dimension case, Eqs. 30 and 31, the setting $\mathbf{X} = \mathbf{X}_{\alpha l} = \mathbf{X}_{\alpha r}$ leads to

$$P_{\alpha l}(A_i) = P_{\alpha r}(A_i) = P(X_1 = t | x_1, x_2, t \in \mathbf{X}; x_1 \leq t_l \leq x_2), \quad (34)$$

Further properties and computation rules for fuzzy probabilities may be derived from traditional probability theory in conjunction with fuzzy set theory.

Representation of Fuzzy Random Variables

The fuzzy probability $\tilde{P}(\underline{A}_i)$ may be computed for each arbitrary set

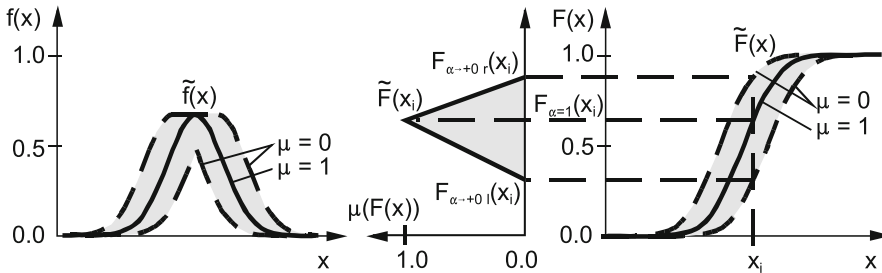
$$\underline{A}_i = \{t = (t_1, \dots, t_k, \dots, t_n) | \underline{x} = \underline{x}_i; \underline{x}, t \in \mathbb{R}^n; t_k < x_k, k = 1 \dots n\} \quad (35)$$

The concept of fuzzy probability according to section “Fuzzy Probability” applied to the sets from Eq. 35 leads to fuzzy probability distribution functions; see Fig. 5. The fuzzy probability distribution function $\tilde{F}(x)$ of the fuzzy random variable $\tilde{\underline{X}}$ on $\mathbf{X} = \mathbb{R}^n$ is obtained as the set of the fuzzy probabilities $\tilde{P}(\underline{A}_i)$ with $\underline{A}_i$ according to Eq. 35 for all $\underline{x}_i \in \mathbf{X}$,

$$\tilde{F}(\underline{x}) = \{\tilde{P}(\underline{A}_i) | \forall \underline{x}_i \in \mathbf{X}\}. \quad (36)$$

It is a fuzzy function. Bounds for the functional values $\tilde{F}(x)$ are specified for each α -level depending on $\underline{x} = \underline{x}_i$ in Eq. 35 and in compliance with Eqs. 20 and 21,

$$F_{\alpha l}(\underline{x} = (x_1, \dots, x_n)) = 1 - \max_{\underline{X}_j \in \underline{X}_x} P(\underline{X}_j = t = (t_1, \dots, t_n) | \underline{x}, t \in \mathbf{X} = \mathbb{R}^n, \exists t_k \geq x_k, 1 \leq k \leq n), \quad (37)$$



Fuzzy Probability Theory, Fig. 5 Fuzzy probability density function $\tilde{f}(\underline{x})$ and fuzzy probability distribution function $\tilde{F}(\underline{x})$ of a continuous fuzzy random variable $\tilde{\underline{X}}$

$$\begin{aligned}
F_{\alpha r}(\underline{x} = (x_1, \dots, x_n)) &= \max_{\underline{X}_j \in \underline{X}_\alpha} P(\underline{X}_j = \\
\underline{t} = (t_1, \dots, t_n) \mid \underline{x}, t \in \underline{X} &= \mathbb{R}^n, \\
t_k < x_k, k = 1, \dots, n).
\end{aligned} \quad (38)$$

For the determination of $F_{\alpha l}(\underline{x})$, the relationship in Eq. 26 is used. If $F_{\alpha l}(\underline{x})$ and $F_{\alpha r}(\underline{x})$ form closed connected intervals $[F_{\alpha l}(\underline{x}), F_{\alpha r}(\underline{x})]$ – see section “Fuzzy Probability” for the conditions – the functional values $\tilde{F}(\underline{x})$ are determined based on Eq. 22,

$$\begin{aligned}
\tilde{F}(\underline{x}) &= \{(F_\alpha(\underline{x}), \mu(F_\alpha(\underline{x}))) \mid F_\alpha(\underline{x}) = [F_{\alpha l}(\underline{x}), F_{\alpha r}(\underline{x})], \\
\mu(F_\alpha(\underline{x})) &= \alpha \forall \alpha \in (0, 1]\}
\end{aligned} \quad (39)$$

In this case, the functional values of the fuzzy probability distribution function $\tilde{F}(\underline{x})$ are continuous and convex fuzzy sets.

In correspondence with Eq. 16, the fuzzy probability distribution function $\tilde{F}(\underline{x})$ of $\tilde{\underline{X}}$ represents the fuzzy set of the probability distribution functions $F_j(\underline{x})$ of all originals $\underline{X}_j$ of $\tilde{\underline{X}}$ with the membership values $\mu(F_j(\underline{x}))$,

$$\begin{aligned}
\tilde{F}(\underline{x}) &= \\
\{(F_j(\underline{x}), \mu(F_j(\underline{x}))) \mid \underline{X}_j \in \tilde{\underline{X}}, \mu(F_j(\underline{x})) &= \mu(\underline{X}_j) \forall j\}.
\end{aligned} \quad (40)$$

Each original $\underline{X}_j$ determines precisely one trajectory $F_j(\underline{x})$ within the bunch $\tilde{F}(\underline{x})$ of weighted functions $F_j(\underline{x}) \in \tilde{F}(\underline{x})$.

In the one-dimensional case and with the restriction to closed random intervals $[X_{\alpha l}, X_{\alpha r}]$ for each α -level, the fuzzy probability distribution function $\tilde{F}(x)$ is determined by

$$F_{\alpha l}(x) = P(X_{\alpha r} = t_r \mid x, t_r \in \mathbf{X}, t_r < x), \quad (41)$$

$$F_{\alpha r}(x) = P(X_{\alpha l} = t_l \mid x, t_l \in \mathbf{X}, t_l < x). \quad (42)$$

In correspondence with traditional probability theory, fuzzy probability density functions $\tilde{f}(\underline{t})$ or $\tilde{f}(\underline{x})$ are defined in association with the $\tilde{F}(\underline{x})$; see Fig. 5. The $\tilde{f}(\underline{t})$ or $\tilde{f}(\underline{x})$ are fuzzy functions

which – in the continuous case with respect to randomness – are integrable for each original $\underline{X}_j$ of $\tilde{\underline{X}}$ and satisfy the relationship

$$F_j(\underline{x}) = \int_{t_1=-\infty}^{t_1=x_1} \dots \int_{t_k=-\infty}^{t_k=x_k} \dots \int_{t_n=-\infty}^{t_n=x_n} f_j(\underline{t}) d\underline{t}, \quad (43)$$

with $\underline{t} = (t_1, \dots, t_n) \in \underline{X}$. For each original $\underline{X}_j$ the integration of the associated trajectory $f_j(\underline{x}) \in \tilde{f}(\underline{x})$ leads to the respective trajectory $F_j(\underline{x}) \in \tilde{F}(\underline{x})$.

For the description of a fuzzy random variable $\tilde{\underline{X}}$, parameters in the form of fuzzy variables $\tilde{p}_t(\tilde{\underline{X}})$ may be used. These fuzzy parameters may represent any type of parameters known from real-valued random variables, such as moments, weighting factors for different distribution types in a compound distribution, or functional parameters of the distribution functions. The fuzzy parameter $\tilde{p}_t(\tilde{\underline{X}})$ of the fuzzy random variable $\tilde{\underline{X}}$ is the fuzzy set of the parameter values $p_t(\underline{X}_j)$ of all originals $\underline{X}_j$ with the membership values $\mu(p_t(\underline{X}_j))$.

$$\begin{aligned}
\tilde{p}_t(\tilde{\underline{X}}) &= \\
\{(p_t(\underline{X}_j), \mu(p_t(\underline{X}_j))) \mid \underline{X}_j \in \tilde{\underline{X}}, \mu(p_t(\underline{X}_j)) &= \mu(\underline{X}_j) \forall j\}.
\end{aligned} \quad (44)$$

For each α -level, bounds are given for the fuzzy parameter $\tilde{p}_t(\tilde{\underline{X}})$ by

$$p_{t, \alpha l}(\tilde{\underline{X}}) = \min_{\underline{X}_j \in \underline{X}_\alpha} [p_t(\underline{X}_j)], \quad (45)$$

$$p_{t, \alpha r}(\tilde{\underline{X}}) = \max_{\underline{X}_j \in \underline{X}_\alpha} [p_t(\underline{X}_j)]. \quad (46)$$

If the fuzzy random variable $\tilde{\underline{X}}$ represents a fuzzy set of continuous real-valued random variables, if all fuzzy realizations $\tilde{x}_j$ of $\tilde{\underline{X}}$ are connected sets, and if the parameter p_t is defined on a continuous scale, then the fuzzy parameter $\tilde{p}_t(\tilde{\underline{X}})$ is determined by its α -level sets

$$p_{t,\alpha}(\tilde{\mathbf{X}}) = [p_{t,\alpha l}(\tilde{\mathbf{X}}), p_{t,\alpha r}(\tilde{\mathbf{X}})], \quad (47)$$

$$\tilde{p}_t(\tilde{\mathbf{X}}) = \{ (p_{t,\alpha}(\tilde{\mathbf{X}}), \mu(p_{t,\alpha}(\tilde{\mathbf{X}}))) \mid \mu(p_{t,\alpha}(\tilde{\mathbf{X}})) = \alpha \forall \alpha \in (0, 1] \}, \quad (48)$$

and represents a continuous and convex fuzzy set.

If a fuzzy random variable $\tilde{\mathbf{X}}$ is described by more than one fuzzy parameter $\tilde{p}_i(\tilde{\mathbf{X}})$, interactive dependencies are generally present between the different fuzzy parameters. If this interaction is neglected, a fuzzy random variable $\tilde{\mathbf{X}}_{hull}$ is obtained, which covers the actual fuzzy random variable $\tilde{\mathbf{X}}$ completely. That is, for all realizations of $\tilde{\mathbf{X}}_{hull}$ and $\tilde{\mathbf{X}}$ the following holds,

$$\tilde{\mathbf{X}}_{hull} \supseteq \tilde{\mathbf{X}} \forall i. \quad (49)$$

Fuzzy parameters and fuzzy probability distribution functions do not enable a unique reproduction of fuzzy realizations based on the above description. But they are sufficient to compute fuzzy probabilities correctly for any events defined according to Eq. 14.

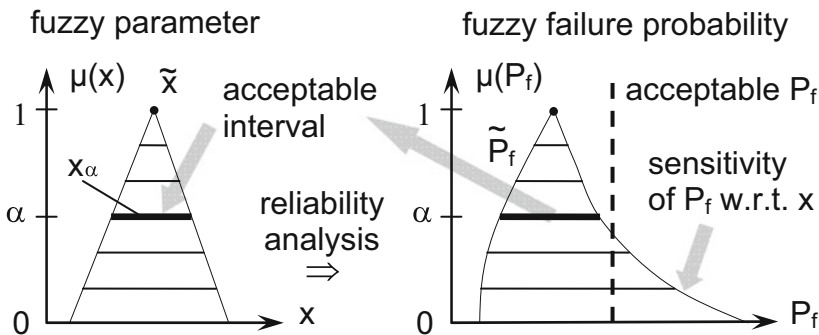
The presented concept of fuzzy probability can be extended to fuzzy random functions and processes.

Application Fields

The practical importance of fuzzy probabilities is based on its transitional position between fuzzy set theory and probability theory. It enables the

consideration of a fuzzy set of probabilistic models, which are variously plausible according to the available information. Aleatory uncertainty and subjective probabilistic information are captured in probabilistic models, and imprecision in the probabilistic model specification is described with fuzzy sets. This preserves uncertainties as probabilistic information and imprecision as set-theoretical information throughout the entire analysis and does not let them migrate into one another. In the case that only fuzzy information is available, the special case of a pure fuzzy analysis appears. On the other hand, if all information can be captured with precisely specified probabilistic models, the result is equal to the traditional probabilistic result.

The practical utilization of these features can be illustrated by interpreting a fuzzy probabilistic analysis as repeated probability-box analysis (Beer et al. 2013a). In the same fashion as a fuzzy number can be described as a family of α -level sets, via α -discretization, a fuzzy set of probabilistic models can be regarded as a set of probability boxes. Figure 6 shows this concept for a fuzzy probabilistic reliability analysis to calculate a fuzzy failure probability $\tilde{P}_f$. In this figure, the fuzzy number $\tilde{x}$ represents a parameter of a fuzzy probabilistic model for the analysis, for example a variance for the distribution of a load. In the analysis, the failure probability P_f of a structure or system is calculated, and the imprecision of $\tilde{x}$ is mapped to this result. Each α -level set x_α of $\tilde{x}$ represents an interval parameter of a



Fuzzy Probability Theory, Fig. 6 Fuzzy probabilistic reliability analysis

probability distribution and thus defines a probability box according to (Ferson and Hajagos 2004). The analysis with this probability box yields an interval for the failure probability associated with the same α -level. Repeating the p-box analysis with several different α then leads to a nested set of α -level sets for the result P_f , which form the fuzzy result $\tilde{P}_f$. This consideration provides a second access to fuzzy probabilities starting from probability boxes to consider various box sizes in a nested fashion in one analysis. Developments to approach fuzzy probabilities from this direction originated from (Kaufman and Gupta 1985) with the notion of hybrid arithmetic; see Ferson and Ginzburg (1995). A fuzzy probabilistic model can thus be formulated in the same manner as a probability box (or interval probability) but provides the additional feature of a nuanced description of the imprecision in the probabilistic model. This is discussed in a geotechnical context in (Beer et al. 2013b), starting from an interval perspective. Discussions on quantification are provided in (Cooper et al. 1995, 1996) and also in (Beer 2009). Interval-valued information in the specification of parameters, distribution types, dependencies, or functional values of a distribution can be implemented including a gradual subjective assessment of the interval sizes. For example, the results from interval estimations on various confidence levels and conflicting statistical test results for various thresholds of rejection probabilities can be used as the basis for a modeling with stepwise changing interval sizes.

In this context, it becomes obvious that the membership function serves only instrumentally to summarize various plausible interval models in one embracing scheme. The interpretation of the membership value μ as epistemic possibility which is sometimes proposed may be useful for ranking purposes but not for making critical decisions. The importance of fuzzy modeling lies in the simultaneous consideration of various magnitudes of imprecision at once in the same analysis. The nuanced features of fuzzy probabilities provide extended insight in technical problems and a workable basis to solve various problems in an elegant and efficient manner.

The largest application field appears as **reliability assessment**, where fuzzy probabilities help to address sensitivities of the failure probability with respect to the probabilistic model choice (Valdebenito et al. 2020; Angelis et al. 2015; Yun et al. 2019). As the tails of the distributions are decisive for the failure probability but can only be determined and justified vaguely based on statistical data and expert knowledge, an analysis with an entire set of plausible probabilistic models and the identification of an associated upper bound for the failure probability are beneficial. This reduces the risk of wrong decisions due to unintentionally optimistic modeling. In the **analysis of sensitivities** of model output, fuzzy probabilities can provide useful new insights with features for systematic and extended investigations (Zhang et al. 2016; Purba et al. 2017; Bi et al. 2019; Schöbi and Sudret 2019). The consideration of fuzzy parameters enables the investigation of sensitivities with respect to changes of the parameters in the entire set of the input, with respect to the magnitude of input imprecision and with respect to the probabilistic model choice. This reveals sensitivities in a global sense over a finite domain rather than in a differential manner. Tolerances given in absolute terms can thereby be translated directly to bounds for model output without extrapolation as required in differential approaches. The advantages are obvious in cases when the model behavior is strongly nonlinear or discontinuous or when derivatives cannot be determined for some reason. Results from a set-theoretical approach are then more robust and reliable. Figure 6 illustrates the identification of sensitivities of the failure probability from a reliability analysis with respect to the imprecision in the probabilistic model specification. Sensitivities of P_f are indicated when the interval size of P_f grows strongly with a moderate increase of the interval size of the input parameters. If this is the case, the membership function of $\tilde{P}_f$ shows outreaching or long and flat tails. A practical consequence would be to pay particular attention to those model options in the input, which cause large intervals of P_f and to further investigate to verify the reasoning for these options and to possibly exclude these critical

cases. Another advantage of the exploration of an entire domain for input parameters is the identification of favorable and less favorable parameter adjustments. This information can be used to collect further information or to perform further analyses systematically in order to identify the causes for sensitivities or to exclude sensitivities by parameter restrictions.

A fuzzy probabilistic analysis also provides interesting features for **design and decision-making**. The analysis can be performed with coarse specifications for design parameters and for probabilistic model parameters in early design stages. The models then allow a stepwise reduction of imprecision as the available information grows over the design process, that is, when design details are specified and implemented. Acceptable intervals for both design parameters and probabilistic model parameters can be determined directly without a repetition of the analysis. Indications are provided in a quantitative manner to collect additional specific information or to apply certain design measures to reduce the input imprecision to an acceptable magnitude. This implies a limitation of imprecision to only those acceptable magnitudes and thus also caters for an optimum economic effort. For example, a minimum sample size or a minimum measurement quality associated with the acceptable magnitude of imprecision can be directly identified. This concept can also be used for the specification of dependencies based on vague information, see Beer et al. (2018).

Further, revealed sensitivities may be taken as a trigger to change the design of the system under consideration to make it more robust. These methods can also be used for the analysis of aged and damaged structures to generate a rough first picture of the structural integrity and to indicate further detailed investigations to an economically reasonable extent – expressed in form of an acceptable magnitude of input imprecision according to some α -level, see Zhang et al. (2016). A more general and comprehensive elaboration on the benefits of using fuzzy probabilities for decision-making is presented in (Aliev et al. 2016). In the area of **model validation and verification**, fuzzy probabilities provide extended

features in two respects. First, they allow the consideration of an entire set of models without prior weighting rather than a single specific one. This represents model uncertainty. Second, imprecision of data can be taken into account without artificial preconditioning of the data, which represents data uncertainty. A particular benefit in this context is the flexibility in matching the model to reality by means of margins and ranges to capture reality instead of approaching it to some limited extent. That is, envelopes are generated for the model rather than approximations. Particular benefits of this concept have been reported for model updating (Bi et al. 2019; Biswal and Ramaswamy 2017; Patelli et al. 2015).

Further developments and applications of fuzzy probabilities and related concepts in an engineering context can be found in the special issues (Beer and Ferson 2013; Beer et al. 2010, 2011, 2012; Möller and Beer 2008; Beer 2015; Augustin et al. 2017) and in the secondary literature to this article.

Future Directions

Fuzzy probability theory provides a powerful key to solving a broad variety of practical problems that defy an appropriate treatment with traditional probabilistic approaches due to imprecision of the information for model specification. Fuzzy probabilities reflect probabilistic uncertainty and non-probabilistic imprecision of the underlying problem simultaneously and separately and provide extended information and decision aids. These features can be utilized in all application fields of traditional probability theory and beyond. Respective developments for utilization can be observed, particularly, in engineering. Potential for further extensive fruitful applications exists, for example, in psychology, economy, financial sciences, medicine, biology, social sciences, and even in law. In all cases, fuzzy probability theory is not considered as a replacement for traditional probability theory but as a beneficial supplement for an appropriate model specification according to the available information in each particular case

and in order to provide extended features for various analyses.

The focus of further developments is seen on both theory and applications. Future theoretical developments may pursue a measure theoretical clarification of the embedding of fuzzy probability theory in the framework of imprecise probabilities under the umbrella of generalized information theory. This is associated with the ambition to unify the variety of available fuzzy probabilistic concepts and to eventually formulate a clear and consistent generalized fuzzy probability theory. Another important issue for future research is the mathematical description and treatment of dependencies within the fuzziness of fuzzy random variables such as non-probabilistic interaction between fuzzy realizations, between fuzzy parameters, and between fuzzy probabilities of certain events. In parallel to theoretical modeling, further effort is worthwhile towards a consistent concept for the statistical evaluation of imprecise data including the analysis of probabilistic and non-probabilistic dependencies of the data.

In view of applications, the further development of fuzzy probabilistic simulation methods is of central importance. This concerns both theory and numerical algorithms for the direct generation of fuzzy random variables – in a parametric and in a nonparametric fashion. Representations and computational procedures for fuzzy random variables must be developed with focus on a high numerical efficiency to enable a solution of large-scale real-world problems. For a spread into practice, it is further essential to elaborate options and potentials for an interpretation and evaluation of fuzzy probabilistic results such as fuzzy mean values or fuzzy failure probabilities. Decision tools need to be developed to translate numerical results into practical conclusions without requiring the user to have extensive specialized knowledge.

A particular challenge is the rapidly growing complexity of our real-world problems, which induces uncertainty and imprecision to a growing extent as well. In this regard, credal networks provide attractive features, which can be understood as an extension of Bayesian networks to

deal with imprecision in probabilities. As Bayesian networks are currently developing their usefulness in engineering, for example in the assessment of reliability and risk in structural and infrastructure engineering, it can be expected that credal networks will also emerge in engineering to deal with cases involving imprecision and fuzzy probabilities, in particular.

In summary, fuzzy probability theory and its further developments significantly contribute to an improved uncertainty quantification according to reality. The most promising potentials for a utilization are seen in nuanced worst-case investigations in terms of probability, in a sensitivity analysis with respect to non-probabilistic imprecision, and in decision-making based on mixed probabilistic/non-probabilistic information.

Bibliography

Primary Literature

- Alefeld G, Herzberger J (1983) Introduction to interval computations. Academic Press, New York
- Aliiev RA, Pedrycz W, Kreinovich V, Huseynov OH (2016) The general theory of decisions. *Inf Sci* 327:125–148. Elsevier Science
- Alvarez DA, Hurtado JE, Ramírez J (2017) Tighter bounds on the probability of failure than those provided by random set theory. *Comput Struct* 189:101–113. Elsevier Science
- Alvarez DA, Uribe F, Hurtado JE (2018) Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory. *Mech Syst Signal Process* 100:782–801. Elsevier Science
- Angelis M, Patelli E, Beer M (2015) Advanced line sampling for efficient robust reliability analysis. *Struct Saf* 52 Part B:170–182. Elsevier Science
- Antonucci A, de Campos CP, Huber D, Zaffalon M (2015) Approximate Credal network updating by linear programming with applications to decision making. *Int J Approx Reasoning* 58:25–38. Elsevier Science
- Auer E, Luther W, Rebner G, Limbourg P (2010) A verified MATLAB toolbox for the Dempster-Shafer theory. In: *Proceedings of the workshop on the theory of belief functions*
- Augustin T, Doria S, Marinacci M (2017) Imprecise probability: theories and applications. *Int J Approx Reasoning* 84:39–40. Elsevier Science
- Bandemer H (1992) Modelling uncertain data. Akademie-Verlag, Berlin
- Bandemer H, Gebhardt A (2000) Bayesian fuzzy kriging. *Fuzzy Sets Syst* 112:405–418

- Bandemer H, Gottwald S (1995) Fuzzy sets, fuzzy logic fuzzy methods with applications. Wiley, Chichester
- Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer Academic Publishers, Dordrecht
- Beer M (2007) Model-free sampling. *Struct Saf* 29:49–65
- Beer M (2009) Engineering quantification of inconsistent information. *Int J Reliab Saf* 3(1–3):174–200
- Beer M (Guest editors), Patelli E (2015) Editorial: engineering analysis with vague and imprecise information. *Struct Saf*, 143. Elsevier Science
- Beer M, Ferson S (eds) (2013) Imprecise probabilities – what can they add to engineering analyses? *Mech Syst Signal Process* 37. Elsevier
- Beer M, Phoon KK, Quek ST (eds) (2010) Special issue: modeling and analysis of rare and imprecise information. *Struct Saf* 32
- Beer M, Muhanna RL, Mullen RL (eds) (2011) Special issue: robust design – coping with hazards risk and uncertainty (Part 1). *Int J Reliab Saf* 5
- Beer M, Muhanna RL, Mullen RL (eds) (2012) Special issue on robust design – coping with hazards risk and uncertainty (Part 2). *Int J Reliab Saf* 6
- Beer M, Ferson S, Kreinovich V (2013a) Imprecise probabilities in engineering analyses. *Mech Syst Signal Process* 37(1–2):4–29. Special Issue: Imprecise probabilities – what can they add to engineering analyses?
- Beer M, Zhang Y, Quek ST, Phoon KK (2013b) Reliability analysis with scarce information: comparing alternative approaches in a geotechnical engineering context. *Struct Saf* 41:1–10
- Beer M, Gong Z, Neumann I, Sriboonchitta S, Kreinovich V (2018) What if we do not know correlations?. In: International econometric conference of Vietnam. Springer, pp 78–85
- Behrendorf J, Broggi M, Beer M (2019) Reliability analysis of networks interconnected with copulas. *ASCE-ASME J Risk Uncertain Eng Syst, Part B: Mech Eng* 5 (4), Article 041006
- Berleant D, Zhang J (2004) Representation and problem solving with distribution envelope determination (denv). *Reliab Eng Syst Saf* 85(1–3):153–168
- Bernardini A, Tonon F (2004) Aggregation of evidence from random and fuzzy sets. Special Issue of ZAMM – Zeitschrift für Angewandte Mathematik und Mechanik 84(10–11):700–709
- Bi S, Broggi M, Wei P, Beer M (2019) The Bhattacharyya distance: enriching the P-box in stochastic sensitivity analysis. *Mech Syst Signal Process* 129:265–281. Elsevier Science
- Biswal S, Ramaswamy A (2017) Finite element model updating of concrete structures based on imprecise probability. *Mech Syst Signal Process* 94:165–179. Elsevier Science
- Bodjanova S (2000) A generalized histogram. *Fuzzy Sets Syst* 116:155–166
- Colubi A, Domínguez-Menchero JS, López-Díaz M, Gil MA (1999) A generalized strong law of large numbers. *Probab Theory Relat Fields* 114:401–417
- Colubi A, Domínguez-Menchero JS, López-Díaz M, Ralescu DA (2001) On the formalization of fuzzy random variables. *Inf Sci* 133:3–6
- Colubi A, Domínguez-Menchero JS, López-Díaz M, Ralescu DA (2002a) A $de[0, 1]$ -representation of random upper semicontinuous functions. *Proc Am Math Soc* 130:3237–3242
- Colubi A, Fernández-García C, Gil MA (2002b) Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. *IEEE Trans Fuzzy Syst* 10:384–390
- Coolen FPA, Coolen-Maturi T (2016) The structure function for system reliability as predictive (imprecise) probability. *Reliab Eng Syst Saf* 154:180–187. Elsevier Science
- Cooper JA, Ferson S, Ginzburg LR (1995) Hybrid processing of stochastic and subjective uncertainty data. Technical report SAND95-2450. Sandia National Laboratories, Albuquerque
- Cooper JA, Ferson S, Ginzburg LR (1996) Hybrid processing of stochastic and subjective uncertainty data. *Risk Anal* 16:785–791
- Couso I, Dubois D (2009) On the variability of the concept of variance for fuzzy random variables. *IEEE Trans Fuzzy Syst* 17:1070–1080
- Couso I, Sanchez L (2008) Higher order models for fuzzy random variables. *Fuzzy Sets Syst* 159(3):237–258
- Crespo LG, Kenny SP, Giesy DP (2013) Reliability analysis of polynomial systems subject to p-box uncertainties. *Mech Syst Signal Process* 37(1–2):121–136. Elsevier Science
- Crespo LG, Colbert BK, Kenny SP, Giesy DP (2019) On the quantification of aleatory and epistemic uncertainty using sliced-normal distributions. *Syst Control Lett* 134, Article 104560. Elsevier Science
- de Cooman G (ed) (2002) The society for imprecise probability: theories and applications. <http://www.sipta.org>
- Degrauwe D, Lombaert G, De Roeck G (2010) Improving interval analysis in finite element calculations by means of affine arithmetic. *Comput Struct* 88(3–4):247–254
- Diamond P (1990) Least squares fitting of compact set-valued data. *J Math Anal Appl* 147:351–362
- Diamond P, Kloeden PE (1994) Metric spaces of fuzzy sets: theory and applications. World Scientific, Singapore/River Edge
- Do DM, Gao W, Song C, Tangaramvong S (2014) Dynamic analysis and reliability assessment of structures with uncertain-but-bounded parameters under stochastic process excitations. *Reliab Eng Syst Saf* 132:46–59. Elsevier Science
- Do DM, Gao W, Song C (2016) Stochastic finite element analysis of structures in the presence of multiple imprecise random field parameters. *Comput Methods Appl Mech Eng* 300:657–688. Elsevier Science
- Dubois D, Prade H (1980) Fuzzy sets and systems theory and applications. Academic Press, New York/London
- Dubois D, Prade H (1985) A review of fuzzy set aggregation connectives. *Inf Sci* 36:85–121

- Dubois D, Prade H (1986) Possibility theory. Plenum Press, New York/London
- Eldred MS, Swiler LP, Tang G (2011) Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation. *Reliab Eng Syst Saf* 96:1092–1113
- Faes M, Moens D (2019) Imprecise random field analysis with parametrized kernel functions. *Mech Syst Signal Process* 134:106334. Elsevier Science
- Fellin W, Lessmann H, Oberguggenberger M, Vieider R (eds) (2005) Analyzing uncertainty in civil engineering. Springer, Berlin/Heidelberg/New York
- Feng Y, Hu L, Shu H (2001) The variance and covariance of fuzzy random variables and their applications. *Fuzzy Sets Syst* 120(3):487–497
- Feng J, Wu D, Gao W, Li G (2017) Uncertainty analysis for structures with hybrid random and interval parameters using mathematical programming approach. *Appl Math Model* 48:208–232. Elsevier Science
- Ferraro MB, Colubi A, Gonzalez-Rodriguez G, Coppi R (2011) A determination coefficient for a linear regression model with imprecise response. *Environmetrics* 22(4):516–529
- Ferson S, Ginzburg L (1995) Hybrid arithmetic. In: Proceedings of the 1995 joint ISUMA/NAFIPS symposium on uncertainty modeling and analysis. IEEE Computer Society Press, Los Alamitos, pp 619–623
- Ferson S, Hajagos JG (2004) Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliab Eng Syst Saf* 85(1–3):135–152
- Fetz T, Oberguggenberger M (2004) Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliab Eng Syst Saf* 85(1–3):73–87
- Fetz T, Oberguggenberger M (2016) Imprecise random variables, random sets, and Monte Carlo simulation. *Int J Approx Reasoning* 78:252–264. Elsevier Science
- Freitag S, Edler P, Kremer K, Meschke G (2020) Multilevel surrogate modeling approach for optimization problems with polymorphic uncertain parameters. *Int J Approx Reasoning* 119:81–91. Elsevier Science
- George-Williams H, Feng G, Coolen FPA, Beer M, Patelli E (2019) Extending the survival signature paradigm to complex systems with non-repairable dependent failures. *J Risk Reliab* 233(4):505–519. SAGE Journals
- Gil MÁ, López-Díaz M, Ralescu DA (2006) Overview on the development of fuzzy random variables. *Fuzzy Sets Syst* 157(19):2546–2557
- González-Rodríguez G, Montenegro M, Colubi A, Gil MÁ (2006) Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. *Fuzzy Sets Syst* 157(19):2608–2613
- Gonzalez-Rodriguez G, Blanco A, Colubi A, Asuncion Lubiano M (2009) Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets Syst* 160(3):357–370
- González-Rodríguez G, Colubi A, Trutschnig W (2009) Simulation of fuzzy random variables. *Inf Sci* 179(5):642–653
- Graf W, Götz M, Kaliske M (2015) Analysis of dynamical processes under consideration of polymorphic uncertainty. *Struct Saf* 52 Part B:194–201. Elsevier Science
- Grzegorzewski P (2000) Testing statistical hypotheses with vague data. *Fuzzy Sets Syst* 112:501–510
- Hall JW, Lawry J (2004) Generation, combination and extension of random set approximations to coherent lower and upper probabilities. *Reliab Eng Syst Saf* 85(1–3):89–101
- Helton JC, Oberkampf WL (eds) (2004) Special issue on alternative representations of epistemic uncertainty. *Reliab Eng Syst Saf* 85
- Helton JC, Johnson JD, Oberkampf WL (2004) An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliab Eng Syst Saf* 85(1–3):39–71
- Hung W-L (2001) Bootstrap method for some estimators based on fuzzy data. *Fuzzy Sets Syst* 119:337–341
- Hurtado JE (2013) Assessment of reliability intervals under input distributions with uncertain parameters. *Probab Eng Mech* 32:80–92. Elsevier Science
- Hwang C-M, Yao J-S (1996) Independent fuzzy random variables and their application. *Fuzzy Sets Syst* 82:335–350
- Jang L-C, Kwon J-S (1998) A uniform strong law of large numbers for partial sum processes of fuzzy random variables indexed by sets. *Fuzzy Sets Syst* 99:97–103
- Joo SY, Kim YK (2001) Kolmogorovs strong law of large numbers for fuzzy random variables. *Fuzzy Sets Syst* 120:499–503
- Kaufman A, Gupta MM (1985) Introduction to fuzzy arithmetic: theory and applications. Van Nostrand Reinhold, New York
- Kim Y-K (2002) Measurability for fuzzy valued functions. *Fuzzy Sets Syst* 129:105–109
- Klement EP (1991) Fuzzy random variables. *Annales Univ Sci Budapest, Sect Comp* 12:143–149
- Klement EP, Puri ML, Ralescu DA (1986) Limit theorems for fuzzy random variables. *Proc R Soc A-Math Phys Eng Sci* 407:171–182
- Klir GJ (2006) Uncertainty and information: foundations of generalized information theory. Wiley-Interscience, Hoboken
- Klir GJ, Folger TA (1988) Fuzzy sets, uncertainty, and information. Prentice Hall, Englewood Cliffs
- Körner R (1997a) Linear models with random fuzzy variables. PhD thesis, Bergakademie Freiberg, Fakultät für Mathematik und Informatik
- Körner R (1997b) On the variance of fuzzy random variables. *Fuzzy Sets Syst* 92:83–93
- Körner R, Näther W (1998) Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Inf Sci* 109:95–118
- Krätschmer V (2001) A unified approach to fuzzy random variables. *Fuzzy Sets Syst* 123:1–9
- Krätschmer V (2002) Limit theorems for fuzzy-random variables. *Fuzzy Sets Syst* 126:253–263

- Krätschmer V (2004) Probability theory in fuzzy sample space. *Metrika* 60:167–189
- Kruse R, Meyer KD (1987) Statistics with Vague data. Reidel, Dordrecht
- Kwakernaak H (1978) Fuzzy random variables – I. Definitions and theorems. *Inf Sci* 15:1–19
- Kwakernaak H (1979) Fuzzy random variables – II. Algorithms and examples for the discrete case. *Inf Sci* 17:253–278
- Li S, Ogura Y, Kreinovich V (2002) Limit theorems and applications of set valued and fuzzy valued random variables. Kluwer Academic Publishers, Dordrecht
- Limbourg P, de Rocquigny E (2010) Uncertainty analysis using evidence theory – confronting level-1 and level-2 approaches with data availability and computational constraints. *Reliab Eng Syst Saf* 95:550–564
- Limbourg P, de Rocquigny E, Andrianov G (2010) Accelerated uncertainty propagation in two-level probabilistic studies under monotony. *Reliab Eng Syst Saf* 95(9): 998–1010
- Lin TY, Yao YY, Zadeh LA (eds) (2002) Data mining, rough sets and granular computing. Physica-Verlag GmbH, Heidelberg
- López-Díaz M, Gil MA (1998) Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *J Stat Plan Inference* 74:11–29
- Matheron G (1975) Random sets and integral geometry. Wiley, New York
- Möller B, Beer M (2004) Fuzzy randomness – uncertainty in civil engineering and computational mechanics. Springer, Berlin
- Möller B, Beer M (eds) (2008) Special issue on uncertainties in structural analysis – their effect on robustness, sensitivity and design. *Comput Struct* 86. Elsevier
- Möller B, Reuter U (2007) Uncertainty forecasting in engineering. Springer, Berlin
- Möller B, Graf W, Beer M (2000) Fuzzy structural analysis using alpha-level optimization. *Comput Mech* 26: 547–565
- Muhanna RL, Mullen RL, Zhang H (2007) Interval finite element as a basis for generalized models of uncertainty in engineering mechanics. *J Reliable Comput* 13(2): 173–194
- Muscolino G, Sofi A (2017) Analysis of structures with random axial stiffness described by imprecise probability density functions. *Comput Struct* 184:1–13. Elsevier Science
- Näther W (2006) Regression with fuzzy random data. *Comput Stat Data Anal* 51:235–252
- Näther W, Körner R (2002) Statistical modelling, analysis and management of fuzzy data, Chapter On the variance of random fuzzy variables. Physica-Verlag, Heidelberg, pp 25–42
- Näther W, Wünsche A (2007) On the conditional variance of fuzzy random variables. *Metrika* 65:109–122
- Oberguggenberger M (2015) Analysis and computation with hybrid random set stochastic models. *Struct Saf* 52 Part B:233–243. Elsevier Science
- Oberkampf WL, Helton JC, Sentz K (2001) Mathematical representation of uncertainty. In: AIAA non-deterministic approaches forum, number AIAA 2001-1645, Seattle
- Patelli E, Alvarez DA, Broggi M, de Angelis M (2015) Uncertainty management in multidisciplinary design of critical safety systems. *J Aerosp Inf Syst* 12 (1):140–169. ARC
- Patelli E, Feng G, Coolen FPA, Coolen-Maturi T (2017) Simulation methods for system reliability using the survival signature. *Reliab Eng Syst Saf* 167:327–337. Elsevier Science
- Pedroni N, Zio E, Ferrario E, Pisanisi A, Couplet M (2013) Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model. *Comput Struct*
- Pedrycz W, Skowron A, Kreinovich V (eds) (2008) Handbook of granular computing. Wiley, New York
- Pivovarov D, Willner K, Steinmann P (2019) On spectral fuzzy? Stochastic FEM for problems involving polymorphic geometrical uncertainties. *Comput Methods Appl Mech Eng* 350:432–461. Elsevier Science
- Purba JH, Tjahjani DTS, Widodo S, Tjahjono H (2017) α -Cut method based importance measure for criticality analysis in fuzzy probability - based fault tree analysis. *Ann Nucl Energy* 110:234–243. Elsevier Science
- Puri ML, Ralescu DA (1983) Differentials of fuzzy functions. *J Math Anal Appl* 91:552–558
- Puri ML, Ralescu D (1986) Fuzzy random variables. *J Math Anal Appl* 114:409–422
- Puri ML, Ralescu DA (1991) Convergence theorem for fuzzy martingales. *J Math Anal Appl* 160:107–122
- Ramos-Guajardo AB, Colubi A, Gonzalez-Rodriguez G, Gil MA (2010) One-sample tests for a generalized Frechet variance of a fuzzy random variable. *Metrika* 71(2):185–202
- Rodríguez-Muñiz L, López-Díaz M, Gil MA (2005) Solving influence diagrams with fuzzy chance and value nodes. *Eur J Oper Res* 167:444–460
- Sadeghi J, Angelis M, Patelli E (2019) Efficient training of interval Neural Networks for imprecise training data. *Neural Netw* 118:338–351. Elsevier Science
- Sadeghi J, de Angelis M, Patelli E (2020) Robust propagation of probability boxes by interval predictor models. *Struct Saf* 82:101889. Elsevier Science
- Samarasooriya VNS, Varshney PK (2000) A fuzzy modeling approach to decision fusion under uncertainty. *Fuzzy Sets Syst* 114:59–69
- Schöbi R, Sudret B (2017) Structural reliability analysis for p-boxes using multi-level meta-models. *Probab Eng Mech* 48:27–38. Elsevier Science
- Schöbi R, Sudret B (2019) Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions. *Reliab Eng Syst Saf* 187:129–141. Elsevier Science
- Schollmeyer G, Augustin T (2015) Statistical modeling under partial identification: distinguishing three types of identification regions in regression analysis with

- interval data. *Int J Approx Reasoning* 56 Part B:224–248. Elsevier Science
- Safer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
- Song Q, Leland RP, Chissom BS (1997) Fuzzy stochastic fuzzy time series and its models. *Fuzzy Sets Syst* 88: 333–341
- Song J, Wei P, Valdebenito M, Bi S, Broggi M, Beer M, Lei Z (2019) Generalization of non-intrusive imprecise stochastic simulation for mixed uncertain variables. *Mech Syst Signal Process*:106316. Elsevier Science
- Song J, Valdebenito M, Wei P, Broggi M, Beer M, Lu Z (2020) Non-intrusive imprecise stochastic simulation by line sampling. *Struct Saf* 84, Article 101936, in press. Elsevier Science
- Stein M, Beer M, Kreinovich V (2013) Bayesian approach for inconsistent information. *Inf Sci* 245:96–111 in press. Special Issue: Statistics with Imperfect Data
- Taheri SM, Behboodan J (2001) A bayesian approach to fuzzy hypotheses testing. *Fuzzy Sets Syst* 123:39–48
- Terán P (2006) On borel measurability and large deviations for fuzzy random variables. *Fuzzy Sets Syst* 157(19): 2558–2568
- Terán P (2007) Probabilistic foundations for measurement modelling with fuzzy random variables. *Fuzzy Sets Syst* 158(9):973–986
- Tolo S, Patelli E, Beer M (2018) An open toolbox for the reduction, inference computation and sensitivity analysis of Credal Networks. *Adv Eng Softw* 115:126–148. Elsevier Science
- Troffaes MCM (2018) Imprecise Monte Carlo simulation and iterative importance sampling for the estimation of lower previsions. *Int J Approx Reasoning* 101:31–48. Elsevier Science
- Trutschnig W (2008) A strong consistency result for fuzzy relative frequencies interpreted as estimator for the fuzzy-valued probability. *Fuzzy Sets Syst* 159(3): 259–269
- Utkin LV (2019) An imprecise extension of SVM-based machine learning models. *Neurocomputing* 331:18–32. Elsevier Science
- Valdebenito MA, Beer M, Jensen HA, Chen J, Wei P (2020) Fuzzy failure probability estimation applying intervening variables. *Struct Saf* 83:101909. Elsevier Science
- Viertl R (1996) Statistical methods for non-precise data. CRC Press, Boca Raton/New York/London/Tokyo
- Viertl R (2011) Statistical methods for fuzzy data. Wiley, Chichester
- Viertl R, Hareter D (2004) Generalized Bayes' theorem for non-precise a-priori distribution. *Metrika* 59:263–273
- Viertl R, Trutschnig W (2006) Fuzzy histograms and fuzzy probability distributions. In: Proceedings of the 11th conference on information processing and management of uncertainty in knowledge-based systems, Paris, 2006. Editions EDK. CD-ROM
- Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman & Hall, London/New York
- Walter G, Aslett LJM, Coolen FPA (2017) Bayesian non-parametric system reliability using sets of priors. *Int J Approx Reasoning* 80:338–351. Elsevier Science
- Wang G, Zhang Y (1992) The theory of fuzzy stochastic processes. *Fuzzy Sets Syst* 51:161–178
- Wang G, Zhong Q (1994) Convergence of sequences of fuzzy random variables and its application. *Fuzzy Sets Syst* 63:187–199
- Wang C, Zhang H, Beer M (2018) Computing tight bounds of structural reliability under imprecise probabilistic information. *Comput Struct*:92–104. Elsevier Science
- Wei G, Di W, Gao K, Chen X, Tin-Loi F (2018) Structural reliability analysis with imprecise random and interval fields. *Appl Math Model* 55:49–67. Elsevier Science
- Wei P, Song J, Bi S, Broggi M, Beer M, Lu Z, Yue Z (2019a) Non-intrusive stochastic analysis with parameterized imprecise probability models: I. Performance estimation. *Mech Syst Signal Process* 124:349–368. Elsevier Science
- Wei P, Song J, Bi S, Broggi M, Beer M, Lu Z, Yue Z (2019b) Non-intrusive stochastic analysis with parameterized imprecise probability models: II. Reliability and rare events analysis. *Mech Syst Signal Process* 126:227–247. Elsevier Science
- Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. *Int J Approx Reasoning* 24(2–3):149–170
- Yager RR (1984) A representation of the probability of fuzzy subsets. *Fuzzy Sets Syst* 13:273–283
- Yun W, Lu Z, Feng K, Jiang X (2019) A novel step-wise AK-MCS method for efficient estimation of fuzzy failure probability under probability inputs and fuzzy state assumption. *Engineering Structures* 183:340–350. Elsevier Science
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1968) Probability measures of fuzzy events. *J Math Anal Appl* 23:421–427
- Zhang J, Shields MD (2018a) The effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets. *Comput Methods Appl Mech Eng* 334:483–506. Elsevier Science
- Zhang J, Shields MD (2018b) On the quantification and efficient propagation of imprecise probabilities resulting from small datasets. *Mech Syst Signal Process* 98:465–483. Elsevier Science
- Zhang M, Beer M, Quek ST, Choo YS (2010) Comparison of uncertainty models in reliability analysis of offshore structures under marine corrosion. *Struct Saf* 32(6): 425–432
- Zhang H, Dai H, Beer M, Wang W (2013) Structural reliability analysis on the basis of small samples: an interval Quasi-Monte Carlo method. *Mech Syst Signal Process* 37(1–2):137–151
- Zhang MQ, Beer M, Koh CG, Jensen HA (2016) Nuanced robustness analysis with limited information. *ASCE-ASME J Risk Uncertain Eng Syst, Part A* 2(3): Civil Eng, B4015001

- Zhong Q, Yue Z, Wang G (1994) On fuzzy random linear programming. *Fuzzy Sets Syst* 65(1):31–49
- Zimmermann H-J (1992) *Fuzzy set theory and its applications*. Kluwer Academic Publishers, Boston/London

Books and Reviews

- Augustin T, Hable R (2010) On the impact of robust statistics on imprecise probability models: a review. *Struct Saf* 32(6):358–365
- Ayyub BM (1998) *Uncertainty modeling and analysis in civil engineering*. CRC Press, Boston/London/New York
- Benavoli A, Zaffalon M (2013) Density-ratio robustness in dynamic state estimation. *Mech Syst Signal Process* 37(1–2):54–75
- Blockley DI (1980) *The nature of structural design and safety*. Ellis Horwood, Chichester
- Blockley D (2013) Analysing uncertainties: towards comparing bayesian and interval probabilities. *Mech Syst Signal Process* 37(1–2):30–42
- Cai K-Y (1996) *Introduction to fuzzy reliability*. Kluwer Academic Publishers, Boston
- Chou KC, Yuan J (1993) Fuzzy-bayesian approach to reliability of existing structures. *J Struct Engineering* 119(11):3276–3290
- Crespo LG, Kenny SP, Giesy DP (2013) Reliability analysis of polynomial systems subject to p-box uncertainties. *Mech Syst Signal Process* 37(1–2):121–136
- Diego A, Alvarez, Jorge E, Hurtado (2014) An efficient method for the estimation of structural reliability intervals with random sets, dependence modeling and uncertain inputs; *Computers and Structures* 142, 54–63
- Elishakoff I (1999) *Whys and Hows in uncertainty modeling probability, fuzziness and anti-optimization*. Springer, Wien/New York
- Faes M, Sadeghi J, Broggi M, de Angelis M, Patelli E, Beer M, Moens D (2019) On the robust estimation of small failure probabilities for strong nonlinear models. *ASCE-ASME J Risk Uncertain Eng Syst, Part B: Mech Eng* 5(4), Article 041007
- Feng Y (2000) Decomposition theorems for fuzzy supermartingales and submartingales. *Fuzzy Sets Syst* 116: 225–235
- Ferson S, Oberkampf WL (2009) Validation of imprecise probability models. *Int J Reliab Saf* 3(1–3):3–22
- Ferson S, Oberkampf WL, Ginzburg L (2008) Model validation and predictive capability for the thermal challenge problem. *Comput Methods Appl Mech Eng* 197: 2408–2430. <http://www.ramas.com/thermval.pdf>
- Fetz T, Oberguggenberger M (2010) Multivariate models of uncertainty: a local random set approach. *Struct Saf* 32(6):417–424
- Gil MA, López-Díaz M (1996) Fundamentals and bayesian analyses of decision problems with fuzzy-valued utilities. *Int J Approx Reason* 15:203–224
- Gil MÁ, Montenegro M, González-Rodríguez G, Colubi A, Casals MR (2006) Bootstrap approach to the multi-sample test of means with imprecise data. *Comput Stat Data Anal* 51:148–162
- Goulet JA, Michel C, Smith IFC (2013) Hybrid probabilities and error-domain structural identification using ambient vibration monitoring. *Mech Syst Signal Process* 37(1–2):199–212
- Grzegorzewski P (2001) Fuzzy sets b defuzzification and randomization. *Fuzzy Sets Syst* 118:437–446
- Grzegorzewski P (2004) Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the hausdorff metric. *Fuzzy Sets Syst* 148(2):319–328
- Hartmann D, Breidt M, Nguyen VV, Stangenberg F, Höhler S, Schweizerhof K, Mattern S, Blankenhorn G, Möller B, Liebscher M (2008) Structural collapse simulation under consideration of uncertainty – fundamental concept and results. *Comput Struct* 86(21–22):2064–2078
- Helton JC, Cooke RM, McKay MD, Saltelli A (eds) (2006) *Special issue: the fourth international conference on sensitivity analysis of model output – SAMO 2004*. *Reliab Eng Syst Saf* 91
- Hirota K (1992) *An introduction to fuzzy logic applications in intelligent systems*. Kluwer international series in engineering and computer science, chapter Probabilistic sets: probabilistic extensions of fuzzy sets, vol 165. Kluwer, Boston, pp 335–354
- Jalal-Kamali A, Kreinovich V (2013) Estimating correlation under interval uncertainty. *Mech Syst Signal Process* 37(1–2):43–53
- Kim JJ, Reda Taha MM, Noh H-C, Ross TJ (2013) Reliability analysis to resolve difficulty in choosing from alternative deflection models of {RC} beams. *Mech Syst Signal Process* 37(1–2):240–252
- Klement EP, Puri ML, Ralescu DA (1984) *Cybernetics and systems research 2, chapter Law of large numbers and central limit theorems for fuzzy random variables*. Elsevier, North-Holland, Amsterdam, pp 525–529
- Kutterer H (2004) Statistical hypothesis tests in case of imprecise data. V Hotine-Marussi symposium on mathematical Geodesy. Springer, Berlin, pp 49–56
- Kutterer H, Neumann I (2011) Recursive least-squares estimation in case of interval observation data. *Int J Reliab Saf* 5(3–4):229–249
- Li S, Ogura Y (2003) A convergence theorem of fuzzy-valued martingales in the extended Hausdorff metric h (inf). *Fuzzy Sets Syst* 135:391–399
- Li S, Ogura Y, Nguyen HT (2001) Gaussian processes and martingales for fuzzy valued random variables with continuous parameter. *Inf Sci* 133:7–21
- Li S, Ogura Y, Proske FN, Puri ML (2003) Central limit theorems for generalized set-valued random variables. *J Math Anal Appl* 285:250–263
- Liu B (2009) *Uncertainty Theory*. Laboratory, Tsinghua University, Beijing
- Louf F, Enjalbert P, Ladeveze P, Romeuf T (2010) On lack-of-knowledge theory in structural mechanics. *Comptes Rendus Mécanique* 338:424–433
- McGill WL, Ayyub BM (2008) A transferable belief model for estimating parameter distributions in structural reliability assessment. *Comput Struct* 86(10):1052–1060

- Mehl CH (2013) P-boxes for cost uncertainty analysis. *Mech Syst Signal Process* 37(1–2):253–263
- Meng Q, Xiaobo Q (2011) A probabilistic quantitative risk assessment model for fire in road tunnels with parameter uncertainty. *Int J Reliab Saf* 5(3–4):285–298
- Möller B, Graf W, Beer M (2003) Safety assessment of structures in view of fuzzy randomness. *Comput Struct* 81:1567–1582
- Möller B, Liebscher M, Schweizerhof K, Mattern S, Blankenhorn G (2008) Structural collapse simulation under consideration of uncertainty – improvement of numerical efficiency. *Comput Struct* 86(19–20):1875–1884
- Möller B, Reuter U (2008) Prediction of uncertain structural responses using fuzzy time series. *Comput Struct* 86(10):1123–1139
- Möller B, Beer M (2008) Engineering computation under uncertainty – capabilities of non-traditional models. *Comput Struct* 86(10):1024–1041
- Möller B, Beer M, Graf W, Sickert J-U (2006) Time-dependent reliability of textile strengthened rc structures under consideration of fuzzy randomness. *Comput Struct* 84(8–9):585–603
- Montenegro M, Casals MR, Lubiano MA, Gil MA (2001) Two-sample hypothesis tests of means of a fuzzy random variable. *Inf Sci* 133:89–100
- Montenegro M, Colubi A, Casals MR, Gil MA (2004) Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* 59:31–49
- Montenegro M, González-Rodríguez G, Gil MA, Colubi A, Casals MR (2004) Soft methodology and random information systems, chapter Introduction to ANOVA with fuzzy random variables. Springer, Berlin, pp 487–494
- Muhanna RL, Mullen RL (eds) (2004) Proceedings of the 1st NSF workshop on reliable engineering computing, Savannah, Georgia. Center for Reliable Engineering Computing, Georgia Tech Savannah, Savannah
- Muhanna RL, Mullen RL (eds) (2006) Proceedings of the 2nd NSF workshop on reliable engineering computing. Center for Reliable Engineering Computing, Georgia Tech Savannah, Savannah
- Mullen RL, Zhang H, Muhanna RL (2012) Structural analysis with probability-boxes. *Int J Reliab Saf* 6(1–3):110–129
- Muscolino G, Sofi A (2013) Bounds for the stationary stochastic response of truss structures with uncertain-but-bounded parameters. *Mech Syst Signal Process* 37(1–2):163–181
- Nasekhian A, Schweiger HF (2011) Random set finite element method application to tunneling. *Int J Reliab Saf* 5(3–4):299–319
- Näther W (2010) Copulas and t-norms: mathematical tools for combining probabilistic and fuzzy information, with application to error propagation and interaction. *Struct Saf* 32(6):366–371. Modeling and analysis of rare and imprecise information
- Negoita VN, Ralescu DA (1987) Simulation, knowledge-based computing and fuzzy-statistics. Van Nostrand Reinhold, New York
- Oberguggenberger M, Schuëller GI, Marti K (eds) (2004) Special issue on application of fuzzy sets and fuzzy logic to engineering problems. 84(10–11), *ZAMM: Zeitschrift für Angewandte Mathematik & Mechanik*
- Oberguggenberger M, Fellin W (2008) Reliability bounds through random sets: nonparametric methods and geotechnical applications. *Comput Struct* 86(10):1093–1101
- Okuda T, Tanaka H, Asai K (1978) A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Inf Control* 38:135–147
- Ochoa CSO, Velasco AA, Kreinovich V (2012) Model fusion under probabilistic and interval uncertainty, with application to earth sciences. *Int J Reliab Saf* 6(1–3):167–187
- Pannier S, Waurick M, Graf W, Kaliske M (2013) Solutions to problems with imprecise data – an engineering perspective to generalized uncertainty models. *Mech Syst Signal Process* 37(1–2):105–120
- Proske FN, Puri ML (2002) Strong law of large numbers for Banach space valued fuzzy random variables. *J Theor Probab* 15:543–551
- Reddy RK, Haldar A (1992) Analysis and management of uncertainty: theory and applications. Chapter A random-fuzzy reliability analysis of engineering systems. North-Holland, Amsterdam, pp 319–329
- Reddy RK, Haldar A (1992) A random-fuzzy analysis of existing structures. *Fuzzy Sets Syst* 48:201–210
- Reid SG (2013) Probabilistic confidence for decisions based on uncertain reliability estimates. *Mech Syst Signal Process* 37(1–2):229–239
- Roberto Rocchetta, Matteo Broggi, Edoardo Patelli (2018) Do we have enough data? Robust reliability via uncertainty quantification; *Applied Mathematical Modelling* 54, 710–721
- Ross TJ (2004) Fuzzy logic with engineering applications, 2nd edn. Wiley, Chichester
- Ross TJ, Booker JM, Parkinson WJ (eds) (2002) Fuzzy logic and probability applications – bridging the gap. SIAM & ASA, Philadelphia/Alexandria
- Sanchez L, Couso I, Palacios AM, Palacios JL (2013) A methodology for exploiting the tolerance for imprecision in genetic fuzzy systems and its application to characterization of rotor blade leading edge materials. *Mech Syst Signal Process* 37(1–2):76–91
- Sankararaman S, Mahadevan S (2013) Distribution type uncertainty due to sparse and imprecise data. *Mech Syst Signal Process* 37(1–2):182–198
- Sankararaman S, Mahadevan S (2011) Model validation under epistemic uncertainty. *Reliab Eng Syst Saf* 96(9):1232–1241
- Serir L, Ramasso E, Nectoux P, Zerhouni N (2013) E2gkpro: An evidential evolving multi-modeling

- approach for system behavior prediction with applications. *Mech Syst Signal Process* 37(1–2):213–228
- Sickert J-U, Freitag S, Graf W (2011) Prediction of uncertain structural behaviour and robust design. *Int J Reliab Saf* 5(3–4):358–377
- Sniady P, Mazur-Sniady K, Sieniawska R, Zukowski S (2013) Fuzzy stochastic elements method. Spectral approach. *Mech Syst Signal Process* 37(1–2):152–162
- Stojakovic M (1994) Fuzzy random variables, expectation, and martingales. *J Math Anal Appl* 184:594–606
- Terán P (2004) Cones and decomposition of sub- and supermartingales. *Fuzzy Sets Syst* 147:465–474
- Tonon F, Bernardini A (1998) A random set approach to the optimization of uncertain structures. *Comput Struct* 68(6):583–600
- Utkin LV, Kozine IO (2010) On new cautious structural reliability models in the framework of imprecise probabilities. *Struct Saf* 32(6):411–416
- Viertl R, Hareter D (2004) Fuzzy information and imprecise probability. *ZAMM – Zeitschrift für Angewandte Mathematik und Mechanik* 84(10–11)
- Wang Y (2013) Generalized fokker-planck equation with generalized interval probability. *Mech Syst Signal Process* 37(1–2):92–104
- Yamauchi Y, Mukaidono M (1999) Interval and paired probabilities for treating uncertain events. *IEICE Trans Inf Syst* E82–D(5):955–961
- Yang X, Liu Y, Zhang Y, Yue Z (2015) Hybrid reliability analysis with both random and probability-box variables. *Acta Mech* 226:1341–1357. Springer
- Yubin L, Zhong Q, Wang G (1997) Fuzzy random reliability of structures based on fuzzy random variables. *Fuzzy Sets Syst* 86:345–355
- Yue Z, Wang G, Fen S (1996) The general theory for response analysis of fuzzy stochastic dynamical systems. *Fuzzy Sets Syst* 83:369–405
- Zadeh L (1985) Is probability theory sufficient for dealing with uncertainty in ai: a negative view. In: *Proceedings of the 1st annual conference on uncertainty in artificial intelligence (UAI-85)*. Elsevier Science, New York, pp 103–116
- Zhang H, Dai H, Beer M, Wang W (2013) Structural reliability analysis on the basis of small samples: an interval quasi-Monte Carlo method; *Mechanical Systems and Signal Processing*, 37(1–2), 137–151

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions

I. Burhan Türkşen

Head Department of Industrial Engineering,
TOBB-ETÜ, (Economics and Technology
University of the Union of Turkish Chambers and
Commodity Exchanges), Ankara, Turkey

Article Outline

Glossary

Definition of the Subject

Introduction

Type 1 Fuzzy System Models of the Past

Future of Fuzzy System Models

Case Study Applications

Experimental Design

Conclusions and Future Directions

Bibliography

Glossary

Z-FRB Zadeh's linguistic fuzzy rule base.

TS-FR Takagi–Sugeno fuzzy rule base.

c^* the number of rules in the rule base.

nv the number of input variables in the system.

$X = (x_1, x_2, \dots, x_{nv})$ input vector.

x_j the input (explanatory variable), for $j = 1, \dots, nv$.

A_{ji} the linguistic label associated with j th input variable of the antecedent in the i th rule.

B_i the consequent linguistic label of the i th rule.

R_i i th rule with membership function $\mu_i(x_j) : x_j \rightarrow [0, 1]$.

A_i multidimensional type 1 fuzzy set to represent the i th antecedent part of the rules defined by the membership function $\mu_i(x) : x \rightarrow [0, 1]$.

$a_i = (a_{i,1}, \dots, a_{i,nv})$ the regression coefficient vector associated with the i th rule.

b_i the scalar associated with the i th rule in the regression equation.

SFF-LSE “Special Fuzzy Functions” that are generated by the Least Squares Estimation.

SFF-SVM “Special Fuzzy Functions” estimated by Support Vector Machines.

The estimate of y_i would be obtained as $Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X$ with SFF-LSE

y the dependent variable, assumed to be a linear function.

β_j $j = 0, 1, \dots, nv$, indicate how a change in one of the independent variables affects the dependent variable.

$X = (x_{j,k} \mid j = 1, \dots, nv, k = 1, \dots, nd)$ the set of observations in a training data set.

m the level of fuzziness, $m = 1.1, \dots, 2.5$.

c the number of clusters, $c = 1, \dots, 10$.

J the objective function to be minimized.

$\|\cdot\|_A$ a norm that specifies a distance-based similarity between the data vector x_k and a fuzzy.

$A = I$ the Euclidean norm.

$A = COV^{-1}$ the Mahalonobis norm.

COV the covariance matrix.

m^*, c^* the optimal pair.

$v_{X|Y,i} = (x_{1,i}, x_{2,i}, \dots, x_{nv,i}, y_i)$ the cluster centers m^*

for $m = m^*$ and each cluster $i = 1, \dots, c^*$.

$v_{X,i} = (x_{1,i}, x_{2,i}, \dots, x_{nv,i})$ the cluster centers of m^*

the “input space” for $m = m^*$ and $c = 1, \dots, c^*$.

$\gamma_{ik}(x_k)$ the normalized membership values of x data sample in the i th cluster, $i = 1, \dots, c^*$.

$\Gamma_i = (\gamma_{ik} \mid i = 1, \dots, c^*; k = 1, \dots, nd)$ the membership values of x in the i th cluster.

X'_i, X''_i, X'''_i potential augmented input matrices in SFF-LSE.

$f(\vec{x}_k) = \hat{y}_k = \langle \vec{w}, \vec{x}_k \rangle + b$ linear Support Vector Regression (SVR) equation.

$l_\epsilon = |y_k - f(\vec{x}_k)|_\epsilon = \max\{0, |y - f(x)| - \epsilon\}$

ϵ -insensitive loss function.

$\vec{w}, b$ the weight vector and bias term.

$c > 0$ the tradeoff between the empirical error and the complexity term.

$\xi_k \geq 0$ and $\xi_k^* \geq 0$ the slack variables.

α_k and α_k^* Lagrange multipliers.

$\mathbf{K}(\vec{x}_{k'}, \vec{x}_k)$ the kernel mapping of the input vectors.

$$\hat{y}_{ik'}^* = \hat{f}(\vec{x}_{ik'}, \alpha_i, \alpha_i^*) = \sum_{k=1}^{nd} (\alpha_{ik} - \alpha_{ik}^*) \mathbf{K}(\vec{x}_{ik'}, \vec{x}_{ik}) + b_i$$

output value of k th data sample in i th cluster with SFF-SVM.

$\tilde{A} = \{(x, (u, f_x(u))) | x \in X, u \in J_x \subseteq [0, 1]\}$ type 2 fuzzy set, A .

$f_x(u) : J_x \rightarrow [0, 1], \forall u \in J_x \subseteq [0, 1], \forall x \in X$ secondary membership function.

$f_x(u) = 1, \forall x \in X, \forall u \in J_x, J_x \subseteq [0, 1]$ interval value type 2 membership function. (Mamdani and Assilian 1981) the domain of the primary membership is discrete and the secondary membership values are fixed to 1. Thus, the proposed method is utilizing discrete interval valued type 2 fuzzy sets in order to represent the linguistic values assigned to each fuzzy variable in each rule. These fuzzy sets can be mathematically defined as follows:

$\tilde{A} = \int_{x \in X} \left[\sum_{u \in J_x} 1/u \right] / x$ Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS)

$x \in X \subseteq \mathfrak{R}, u \in J_x = \{J_{xr}\}, r = 1, \dots, NM$ a system variable in continuous domain.

u the primary membership value with discrete domain.

J_{xr} r th primary membership value associated with x .

Definition of the Subject

Fuzzy System Modeling (FSM) is one of the most prominent tools that can be used to identify the behavior of highly nonlinear systems with uncertainty. In the past, FSM techniques utilized type 1 fuzzy sets in order to capture the uncertainty in the system. However, since type 1 fuzzy sets express the belongingness of a crisp value x' of an input variable x in a fuzzy set A by a crisp membership value $\mu_A(x')$, they cannot fully capture the uncertainties associated with higher order imprecision in identifying membership functions. In the future, we are likely to observe higher types

of fuzzy sets, such as type 2 fuzzy sets. The use of type 2 fuzzy sets and linguistic logical connectives drew a considerable amount of attention in the realm of fuzzy system modeling in the last two decades. In this paper, we first review type 1 fuzzy system models known as Zadeh, Takagi–Sugeno and Türkşen models; then we review potentially future realizations of type 2 fuzzy systems again under the headings of Zadeh and Takagi–Sugeno and Türkşen fuzzy system models, in contrast to type 1 fuzzy system models. Type 2 fuzzy system models have a higher predictive power. One of the essential problems of type 2 fuzzy system models is computational complexity. In data-driven fuzzy system modeling methods discussed here, Fuzzy C-Means (FCM) clustering algorithm is used in order to identify the system structure.

Introduction

Fuzzy system models are the most successful models to handle uncertainties in decision-making. The major advantages of fuzzy system models are their robustness and transparency. Fuzzy system modeling achieves robustness by using fuzzy sets which incorporates imprecision in system models. In addition, unlike some other system models, such as neural networks, the fuzzy system models are highly descriptive, i.e., transparent.

In the last two decades, researchers proposed several data driven type 1 fuzzy system modeling approaches that can extract the hidden rules of a system behavior automatically by using historical data. The system modeling methods, proposed by Nakanishi et al. (1993), Takagi–Sugeno (1985), Sugeno and Yasukawa (1993), Emami et al. (1998), are among the most notable ones. Since these methods utilize only the historical data, i.e., since they do not require expert knowledge, they are strictly data-driven modeling techniques. Thus, in addition to being robust and transparent, these system-modeling techniques can identify system model structure objectively for a given performance measure.

In these traditional fuzzy system models of the past, the structure is characterized by type 1 fuzzy

sets. Type 1 fuzzy sets, defined on a universe of discourse, maps an element onto a precise number in the unit interval $[0, 1]$. This conflicts with the basic philosophy of fuzzy set and logic theory. In the future, it is expected, fuzzy system models will be formed with higher order fuzzy sets, such as type 2 fuzzy sets, which was first proposed by Zadeh (1975). They will be used more and more in order to capture the uncertainty associated with membership functions. A type 2 fuzzy set can be informally defined as a fuzzy set that is characterized by a fuzzy membership function, i.e., membership values also are in the unit interval $[0, 1]$ in the computational level.

In this paper, we propose to review first the type 1 fuzzy system models as the historically significant but successful modeling activity of the past in the domain of fuzzy control system problems. Next we review the type 2 fuzzy system models as the potential future modeling activity in the domain of fuzzy decision support system problems.

In an historical sense, Zadeh (1975), and Takagi–Sugeno (1985) versions of type 1 fuzzy system models are typically basic fuzzy rule base models. In contrast, type 1 “special fuzzy function” models, recently proposed by Türkşen (2008), are alternate models to fuzzy rule base models. They give better predictions in comparison to type 1 fuzzy rule base models (Celikyilmaz and Türkşen 2007; Türkşen 2008). Furthermore, fuzzy rule base models, in general, can not capture the interactive nature of a problem space due to projection deficiency. Whereas fuzzy function models are able to capture the interactions of all variables since they are not subject to projection deficiencies.

For future developments, currently there are mainly two essentially different schools of thought in type 2 fuzzy system modeling research. The first one is based on the properties of the linguistic connectives which causes the generation of interval valued type 2 fuzzy sets. Türkşen (1986, 1992, 1995, 2002) mathematically showed that Fuzzy Conjunctive Canonical Forms (FCCF) and Fuzzy Disjunctive Canonical Forms (FDCF) are no longer equivalent to each other for the 16 basic linguistic

expressions formed with linguistic connectives “AND”, “OR”, etc. Furthermore, it is shown that FCCF contains FDCF for certain families of t-norms and t-co-norms. Zimmerman and Zysno (1980) empirically showed that linguistic connectives were characterized differently by different people and Türkşen (1995) provided the ground-work for this with the Interval Valued type 2 fuzzy sets. These type 2 system models represent uncertainties generated by different linguistic connectives that combine type 1 membership functions. Such combinations generate their FDCF and FCCF expressions and hence identify a particular sort of Interval Valued type 2 representations of systems.

In the second school of thought in type 2 fuzzy system modeling, traditional connectives, i.e., t-norms and co-norms, are utilized directly with the assumption that a t-norm directly corresponds to a linguistic “AND” in its FDCF and a t-co-norm directly corresponds to an “OR” in its FCCF while ignoring FCCF of “AND” and FDCF of “OR”. But, each variable is represented by using a type 2 fuzzy set at the beginning of computations. In a series of papers Mendel, Karnik and Lian (Karnik and Mendel 1998b; Karnik et al. 1999; Karnik and Mendel 2000) extended traditional type 1 system models to type 2 fuzzy system models and proposed inference methods in order to process type 2 fuzzy sets. These studies were explained thoroughly by Mendel in (2001). Several other researchers such as, Starczewski and Rutkowski (2002), John and C. Czarnecki (1998, 1999), and Chen and Kawase (2000) worked on type 2 fuzzy system models and inference methods.

In general the main inhibitor of the use of the type 2 fuzzy system models is the computational complexity of the inference mechanism that is used to infer a model output by using type 2 fuzzy system models for a given input data vector. Liang and Mendel (2000) proposed Interval-Valued (IV) type 2 fuzzy system models in place of full type 2 fuzzy system models together with inference methods to remedy this problem. Thus they use a simplified version of type 2 fuzzy sets, namely, Interval-Valued type 2 fuzzy sets (IVT2FS) rather than full type 2 fuzzy sets (FT2FS).

Recently, Uncu and Türkşen (2007b) proposed discrete “interval valued type 2 fuzzy sets (DIVT2FS)” which are generated by a variation of the level of fuzziness around fixed cluster centers in applications of FCM. The proposed model representation enables us to identify a computationally efficient inference mechanism.

The rest of this paper is organized as follows: the basic notation, terminology and the three well-known type 1 fuzzy system model structures will be briefly reviewed in section “Introduction” with an emphasis on the more recent “special fuzzy functions”. Then the extensions of these well-known fuzzy system modeling structures in type 2 formation will be discussed for potential future studies in section “Future of Fuzzy System Models”. Finally, the conclusions will be drawn and the future research directions will be provided in section “Conclusions and Future Directions”.

Type 1 Fuzzy System Models of the Past

In general fuzzy system models identify an underlying relationship between input and output variables of a system. In this paper, we deal with Multi-Input Single Output (MISO) version fuzzy system models. Generally fuzzy system models represent relationships between the input and output variables as a collection of: (a) either if-then rules that utilize linguistic labels (i.e., fuzzy sets) or (b) “special fuzzy functions” that takes on membership values and/or their transformations as well as selected input variables as their arguments.

Type 1 Fuzzy Rulebases

In general, the fuzzy rule base structure can be written as follows:

$$R : \text{ALSO } \left(\text{IF antecedent}_i \text{ THEN consequent}_i \right), \quad (1)$$

where c^* is the number of rules in the rule base. There are several well known fuzzy rule base structures which mainly differ in the representation of their consequents. If the consequent is represented with fuzzy sets then the rule base can be categorized as the Zadeh Fuzzy Rule base, Z-FRB (Zadeh 1975) Whereas, if the

consequent is represented with a linear equation of input variables, then the rule base structure is known as Takagi–Sugeno Fuzzy Rulebase (TS-FRB) structure (Takagi and Sugeno 1985). The Z-FRB and TS-FRB structures can be formalized as follows:

Let nv be the number of input variables in the system. Then, the multidimensional antecedent, X , can be defined as $X = (x_1, x_2, \dots, x_{nv})$, where x_j is the j th input variable of the antecedent in the domain of X . X , can be defined as $X = X_1 \times X_2 \times \dots \times X_{nv}$, where $X_j \subseteq \mathfrak{R}$ is the domain of variable x_j . Similarly, the domain of the output variable, will be denoted as $Y \subseteq \mathfrak{R}$. Then, the i th rule, r_i , and rulebase, R , in Z-FR structure can be defined as:

$$R_i : \text{IF } \bigwedge_{j=1}^{NV} (x_j \in X_j \text{ isr } A_{ji}) \text{ THEN } y \in Y \text{ isr } B_i, \quad (2)$$

$$\forall i = 1, \dots, c^*$$

$$R : \text{ALSO } \left(\text{IF } \bigwedge_{j=1}^{NV} (x_j \in X_j \text{ isr } A_{ji}) \text{ THEN } y \in Y \text{ isr } B_i \right), \quad (3)$$

where A_{ji} is the linguistic label associated with j th input variable of the antecedent in the i th rule, R_i , with membership function $\mu_i(x_j) : x_j \rightarrow [0, 1]$ and similarly B_i is the consequent linguistic label of the i th rule with membership function $\mu_i(y) : y \rightarrow [0, 1]$, and c^* is the number of rules in the model. The above structure assumes non-interactivity between input variables because the membership functions of every A_{ji} is obtained by the projection of $nv + 1$ dimensional internal system representation. In order to eliminate the non-interactivity assumption, Delgado et al. (1997), Babuska and Verbruggen (1997), and Uncu and Türkşen (2007b) used multidimensional type 1 fuzzy sets to represent the antecedent part of the rules. Hence, the Z-FRB structure can be expressed as follows:

$$R : \text{ALSO } \left(\text{IF } x \in X \text{ isr } A_i \text{ THEN } y \in Y \text{ isr } B_i \right), \quad (4)$$

where the multidimensional antecedent fuzzy set of i th rule is defined as $\mu_i(x) : x \rightarrow [0, 1]$. The other

well-known fuzzy rulebase structures, namely Takagi–Sugeno (TS-FRB) fuzzy rulebase structure, can be expressed, respectively, as follows:

$$\text{ALSO}_{i=1}^{c^*} (\text{IF antecedent}_i \text{ THEN } y_i = a_i x^T + b_i), \quad (5)$$

where, $a_i = (a_{i,1}, \dots, a_{i,m})$ is the regression coefficient vector associated with the i th rule, b_i is the scalar associated with the i th rule.

Type 1 “Special Fuzzy Functions” of the Recent Past

There are at least two ways to form special fuzzy functions: (i) “Special Fuzzy Functions” that are generated by the Least Squares Estimation, SFF-LSE, of Türkşen (2008) and (ii) “Special Fuzzy Functions” estimated by Support Vector Machines, SFF-SVM’s, of Çelikyılmaz and Türkşen (2007). These “Special Fuzzy Functions” are structurally different from (1) Zadeh’s (1975) linguistic fuzzy rule bases (Z-FRB), (2) Takagi–Sugeno fuzzy rule base TS-FR (Takagi and Sugeno 1985) (3) “Fuzzy Regression” models of Tanaka et al. (1982, 1995) and its variations, and (5) Hathaway and Bezdek (1993) model. Because the proposed “Special Fuzzy Functions” introduces membership values and their transformations as new input variables in addition to the original scalar input variables in function estimations. In particular, these “Special Fuzzy Functions” are structurally different from (1) “Fuzzy Regression” models of Tanaka et al. (1982, 1995), and its variations, and (2) Hathaway and Bezdek (1993) models. This is why we call them “Special Fuzzy Functions”.

It ought to be noted that the introduction of membership values and their transformations as new input variables are acceptable since the membership values are obtained from FCM which is a highly nonlinear transformation of the original scalar input variable. Hence there is no co-linearity and thus they can be included in the proposed LSE and SVM structure. For this purpose, first one executes a fuzzy clustering algorithm such as FCM with original selected input

variables after an execution of a feature selection algorithm; and then determines (local) optimum number of fuzzy clusters and hence the associated membership values. Then a special fuzzy function to represent each fuzzy cluster, i.e., fuzzy rule, separately can be identified. Thus there are as many fuzzy functions as there are fuzzy clusters similar to Hathaway and Bezdek’s (1993) Fuzzy C-Regression model (FCRM). But Hathaway and Bezdek use membership values as the weights to be used in the estimation of the functions using weighted least squares algorithm. FCRM updates the membership values as the similarity measure by using estimation error from these functions.

“Special Fuzzy Functions”, SFF, are estimated after one generates membership values of each cluster from FCM algorithm. Therefore it is structurally a new and unique approach for the determination of fuzzy system models in place of fuzzy rule bases. This the reason we call them “Special Fuzzy Functions”, SFF. These “Special Fuzzy Functions” represent fuzzy rule bases in functional form and structure. When the relationship between input variables and the output variable of the system can be linearly explained in the original dimension space of the data, it is quite reasonable and faster to estimate such “Special Fuzzy Functions” using least squares estimation. When this relationship is more complex and there needs to be a non-linear transformation of the original input variables, it is better to map the input dataset into a higher dimensional space, e.g., a *hyper-space*, where the input dimension is large (maximum n). One of the powerful methods to find these “Special Fuzzy Functions” which define a linear relation between input and output variables in the higher dimension, but a non-linear relationship in the original dimension is the support vector machines which was first proposed by Vapnik (1998). For the regression case, support vector machines for regression algorithm can be applied to find these “Special Fuzzy Functions”. Hence, in the next sections, we are going to specify the details of these “Special Fuzzy Functions” estimated using the Least Squares, LSE, and Support Vector machine for Regression, SVR, algorithms.

It is to be noted for the sake of emphasis that the estimated parameters of the inputs are not fuzzy sets in our proposed approach. It should be recalled that membership values and their transformations are augmented into the original selected input set as new and additional variables. In our experience, it is found that this approach is most suitable for those analysts who are familiar with a function estimation technology, e.g., the least squares technology, support vector machines, ridge regression, etc. They only need to develop an understanding of some fuzzy clustering algorithm without studying many aspects of fuzzy theory. All they have to understand is the notion of membership values and how they can be obtained from a fuzzy clustering algorithm such as FCM and/or its variations in addition to their usual background knowledge of a function estimation technique, e.g., LSE, or SVR, etc.

Thus this is a novel approach in order to provide an easy entry into fuzzy system modeling for mathematicians and statisticians who are working in industry and for other novices. For this purpose, we present next our generalization of the LSE algorithm, which includes membership values and their transformations in addition to the original scalar input variables.

Special Fuzzy Functions with LSE (SFF-LSE) Method

In Ordinary LSE (OLSE) method, the dependent variable, y , is assumed to be a linear function of one or more independent, input, variables, x , plus an error component as follows:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{nv} x_{nv} + \varepsilon,$$

where y is the dependent output, x_j 's are the inputs (explanatory variables), for $j = 1, \dots, nv$, nv is the number of selected inputs and ε is the independent error term which is typically assumed to be normally distributed. The goal of the least squares method is to obtain estimates of the unknown parameters, β_j 's, $j = 0, 1, \dots, nv$, which indicate how a change in one of the independent variables affects the dependent variable as follows:

$$\beta = (X'X)^{-1}X'Y, \quad (6)$$

where $\beta = (\beta_0, \beta_1, \dots, \beta_{nv})$.

The proposed generalization of OLSE as **SFF-LSE**, requires that a fuzzy clustering algorithm, such as FCM (Bezdek 1973), be available to determine the interactive (joint) membership values of input-output variables in each of the fuzzy clusters that can be identified for a given training data set.

Let (x_k, y_k) , $k = 1, \dots, nd$, be the set of observations in a training data set, such that

$$X = (x_{j,k} \mid j = 1, \dots, nv, k = 1, \dots, nd). \quad (7)$$

First, one determines the optimal (m^*, c^*) pair for a particular performance measure, i.e., a cluster validity index, with an iterative search and an application of FCM algorithm, where m is the level of fuzziness (in our experiments, we usually take $m = 1.1, \dots, 2.5$), and c is the number of clusters (in our experiments, we usually take $c = 2, \dots, 10$). The well known FCM algorithm can be stated as follows:

$$\begin{aligned} \min J(U, V) &= \sum_{k=1}^{nd} \sum_{i=1}^c (u_{ik})^m (\|x_k - v_i\|_A) \\ \text{s.t. } 0 &\leq u_{ik} \leq 1, \quad \forall i, k \\ \sum_{i=1}^c u_{ik} &= 1, \quad \forall k \\ 0 &\leq \sum_{k=1}^{nd} u_{ik} \leq nd, \quad \forall i, \end{aligned}$$

where J is objective function to be minimized, $\|\cdot\|_A$ is a norm that specifies a distance-based similarity between the data vector x_k and a fuzzy cluster center v_i . In particular, $A = I$ is the Euclidean norm and $A = COV^{-1}$ is the Mahalanobis norm, etc., where COV is the covariance matrix.

The optimal pair, (m^*, c^*) , can be determined with a user defined cluster validity index, partition entropy or partition coefficient (Bezdek 1973). Another alternative of selecting the optimum pair would be running the overall SFF-LSE model for every (m, c) pair specified by the user and determining the optimal pair from the training

RMSE values of each model. The following definitions adopt the idea of using a user defined cluster validity index for the determination of the optimal pair. The experiments in this paper follow the second alternative.

Once the optimal pair (m^*, c^*) is determined with the application of FCM algorithm, one next identifies the cluster centers for $m = m^*$ and each cluster $i = 1, \dots, c^*$ as:

$$v_{X|Y,i} = (x_{1,i}, x_{2,i}, \dots, x_{nv,i}, y_i). \quad (8)$$

From this, we identify the cluster centers of the “input space” for $m = m^*$ and $c = 1, \dots, c^*$ as:

$$v_{X,i} = (x_{1,i}, x_{2,i}, \dots, x_{nv,i}). \quad (9)$$

Next, one computes the normalized membership values of each data sample in the training data set with the use of the cluster center values determined in the previous step. There are generally two steps in these calculations:

- (a) first we determine the (local) optimum membership values u_{ik} ’s and then determine μ_{ik} ’s that are above an α -cut in order to eliminate harmonics generated by FCM as:

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{\|x_k - v_{X,i}\|}{\|x_k - v_{X,j}\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad (10)$$

$$\mu_{ik} = \{u_{ik} \geq \alpha\},$$

where μ_{ik} denotes the membership value of the k th vector, $k = 1, \dots, nd$, in the i th rule, $i = 1, \dots, c^*$ and x_k denotes the k th vector.

- (b) next, we normalize them as:

$$\gamma_{ik}(x_k) = \frac{\mu_{ik}(x_k)}{\sum_{i'=1}^c \mu_{i'k}(x_k)}, \quad (11)$$

where $\gamma_{ik}(x_k)$ ’s are the normalized membership values of x data sample in the i th cluster, $i = 1,$

$\dots, c^*$, which in turn indicate the membership values that will constitute as a new input variable in our proposed scheme of function identification for the representation of i th cluster. Let $\Gamma_i = (\gamma_{ik} \mid i = 1, \dots, c^*; k = 1, \dots, nd)$ be the membership values of x, a data sample, in the i th cluster, i.e., i th rule.

Next we determine a new augmented input matrix of x for each of the clusters, which could take on several forms depending on which *transformation* of membership values we want to or need to include in our system structure identification for our intended system analyses. Examples of possible augmented input matrices are:

$$\begin{aligned} X'_i &= [1, \Gamma_i, X], \text{ or} \\ X''_i &= [1, \Gamma_i^2, X], \text{ or} \\ X'''_i &= [1, \Gamma_i^2, \Gamma_i^m, \exp(\Gamma_i), X], \text{ etc.,} \end{aligned} \quad (12)$$

where X'_i, X''_i, X'''_i are potential augmented input matrices to be used in least squares estimation of a new system structure identification and $\Gamma_i = (\gamma_{ik} \mid i = 1, \dots, c^*; k = 1, \dots, nd)$. The choice amongst X'_i, X''_i, X'''_i depends on whether we want to or need to include just the membership values or some of their transformations as new input variables in order to obtain the best representation of a system behavior. A new augmented input matrix having a single input variable in the original input space when only membership values themselves are augmented to the dataset, i.e., X'_i may look like this:

$$X'_i = [1, \Gamma_i, X] = \begin{bmatrix} 1 & \gamma_{i,1} & x_{i,1} \\ \vdots & \vdots & \vdots \\ 1 & \gamma_{i,nd} & x_{i,nd} \end{bmatrix}$$

Up to this point, in the proposed system modeling approach, we have defined how the augmented input matrix for each cluster could be formed with the output of an FCM algorithm. Both the proposed SFF-LSE and SFF-SVM approaches implement these steps. From this point forward, the estimation of “Special Fuzzy Functions” takes place for each cluster, where one can implement any function estimation

methodology, e.g., LSE or SMV. Different approaches are followed in the estimation of “Special Fuzzy Functions” using an augmented matrix. Here we continue to specify the SFF-LSE models. In the next section the SFF-SVM models will be introduced.

Thus the function of a single input single output model, which includes only the membership values as the additional input variable, $Y_i = \beta_{i0} + \beta_{i1}\Gamma_i + \beta_{i2}X$, that represents the i th rule corresponding to the i th interactive (joint) cluster in (Y_i, Γ_i, X) space, would be estimated with SFF-LSE approach as follows:

$$\beta_i^* = (X_i'^T X_i')^{-1} (X_i'^T Y_i), \quad (13)$$

where $\beta_i^* = (\beta_{i0}^*, \beta_{i1}^*, \beta_{i2}^*)$ and $X_i' = [1, \Gamma_i, X]$, provided the inverse of covariance, $(X_i'^T X_i')^{-1}$, exists. The estimate of y_i would be obtained as:

$$Y_i^* = \beta_{i0}^* + \beta_{i1}^* \Gamma_i + \beta_{i2}^* X. \quad (14)$$

The single output value is calculated using each output value, one from each cluster, and weighting them with their corresponding membership values as follows:

$$Y_i^* = \frac{\sum_{i=1}^{c^*} \gamma_i Y_i^*}{\sum_{i=1}^{c^*} \gamma_i}. \quad (15)$$

Within the proposed framework, the general form of the shape of a cluster for the case of a single input variable X_j and for the i th cluster can be conceptually captured, in a stylistic, imaginary manner, say, by a second order (cone) function when one introduces the square of membership values into the augmented input matrix in the space of $[U \times X \times Y]$ which can be illustrated with a prototype shown in Fig. 1.

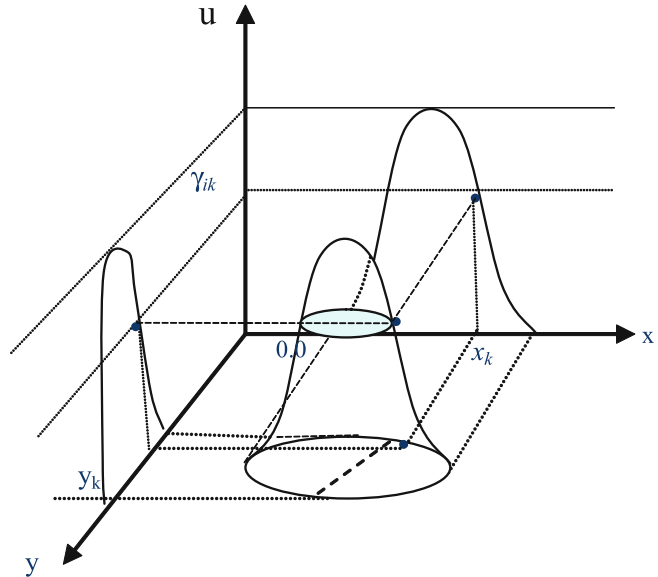
In a number of real life case studies, we have in fact found out that generally some second order or exponential function provide a good approximation from amongst some 20 alternatives we have experimented with in the past.

Before we specify the details of the proposed SFF-SVM method, we first review briefly the background of the support vector machines for regression algorithm.

Support Vector Machines for Regression

Support Vector Machine, SVM, is a data-mining tool to build a model of a given system. The

**Fuzzy System Models
Evolution from Fuzzy
Rulebases to Fuzzy
Functions, Fig. 1** A fuzzy
cluster in $[U \times X \times Y]$ space



foundations of SVM have been developed by Vapnik (1998). SVM is a type of optimization technique in which prediction error and model complexity are simultaneously minimized. Let the training samples be denoted as:

$$X = \left\{ \left(\vec{x}_k, \vec{y}_k \right) \mid k = 1, \dots, nd \right\}, \quad (16)$$

where X denotes the space of input-output patterns $\vec{x}_k = (x_{jk} \mid j = 1, \dots, j_{nv}, k = 1, \dots, k_{nd})$ represents each input data vector, and y_k is the output value of the k th, data vector in the dataset. Support vector machines are used to solve classification problems as well as regression models where the output variable is scalar. In linear Support Vector Regression (SVR), the aim is to find a pair $(\vec{w}, b)$, where $\vec{w}$ is the weight vector, and b is the bias term in this regression equation, such that the value of the point, y_k , can be predicted according to a real-valued function:

$$f(\vec{x}_k) = \hat{y}_k = \langle \vec{w}, \vec{x}_k \rangle + b, \quad (17)$$

where $\langle \cdot \rangle$ is the dot product representation. The goal is to find a function, that has at most ε deviation from the actually obtained targets, y_k , for all the training data. This concept of ε -insensitive loss function, l_ε , was first introduced by Vapnik (Mendel 2001) as follows:

$$\begin{aligned} l_\varepsilon &= \left| y_k - f(\vec{x}_k) \right|_\varepsilon \\ &= \max\{0, |y - f(x)| - \varepsilon\}. \end{aligned} \quad (18)$$

The loss function does not penalize errors below some error, $\varepsilon \geq 0$. Thus the goal of learning is to find a function with a small risk on test samples. This would mean good generalization. This type of SVR is called the ε -insensitive which embodies the Structural Risk Minimization (SRM) as displayed as follows:

$$R_{\text{exp}}[f] \leq R_{\text{emp}}[f] + R_{\text{complexity}}[f]. \quad (19)$$

In **SRM of SVM**, the aim is not only minimize the empirical risk from training samples, R_{emp} , but also find a simple function to minimize the complexity of the model, $R_{\text{complexity}}$. The more flat the functions are, the less complex they would be, in other words they would get simpler, and therefore they would be closer to the linear functions. The more flat function, the smaller the complexity term of the expected risk in SRM, R_{exp} , gets and the smaller would be the weight vector. In support vector regression the complexity term is expressed as weights assigned to all points in the training sample. In order to ensure that the weight vector is small, Euclidean Norm, i.e., $\|\vec{w}\|^2$ is used. In mathematical terms the objective function of SVM for regression with two conditions can be stated as follows.

$$\begin{aligned} \min_{\vec{w}, b, \xi_k, \xi_k^*} \quad & \varsigma(\vec{w}, \xi, \xi^*) = \frac{1}{2} \|\vec{w}\|^2 + \frac{C}{nd} \sum_{k=1}^{nd} (\xi_k + \xi_k^*) \\ \text{subject to} \quad & y_k - \langle \vec{w}, \vec{x}_k \rangle - b \leq \varepsilon + \xi_k \quad \langle \vec{w}, \vec{x}_k \rangle + b - y_k \leq \varepsilon + \xi_k^* \quad \xi_k, \xi_k^* \geq 0, \end{aligned} \quad (20)$$

where ε represents the ε insensitive value which does not penalize the points whose estimated deviations are lesser/greater than ε , and $\vec{w}, b$ unknown values that represent the weight vector and bias term, respectively. The $c > 0$ determines the trade off between the empirical error and the complexity term. The slack variables $\xi_k \geq$

0 and $\xi_k^* \geq 0$ are introduced to the model to soften the optimization problem in order to prevent infeasible solutions. The assumption in (20) is that, it is possible to find such a function that approximates all pairs $(\vec{x}_k, y_k)$ with ε precision. The optimization problem is a convex quadratic program, which can be solved by using the well-

known Lagrange multiplier method. Therefore by introducing Lagrange multipliers α_i and β_i , one can construct Lagrange function, and the solution to the optimization the orem is given by the saddle point of the Lagrange function using the Karush–Kuhn–Tucker theorem (Vapnik 1998) where the primal model is translated into dual quadratic programming problem as follows:

$$\begin{aligned} \max_{\alpha, \alpha^*} &= \frac{1}{2} \sum_{k, k'=1}^{nd} (\alpha_k - \alpha_k^*) (\alpha_{k'} - \alpha_{k'}^*) \langle \vec{x}_k, \vec{x}_{k'} \rangle \\ &\quad - \varepsilon \sum_{k=1}^{nd} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) y_k \\ \text{subject to} &\quad \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) = 0 \\ &\quad \alpha_k, \alpha_k^* \in [0, C]. \end{aligned} \quad (21)$$

In model (21), we search for the parameters α_k and α_k^* , which are Lagrange multipliers. The weight vector can now be explained using the Lagrange multipliers as:

$$\vec{w} = \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \vec{x}_k, \quad (22)$$

where, $\vec{x}_k$ is the k th observation and α_k and α_k^* are the Lagrange multipliers for the k th observation, and the estimation function of a vector is given as follows:

$$\begin{aligned} f(\vec{x}_{k'}) &= \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \langle \vec{x}_{k'}, \vec{x}_k \rangle + b \\ &= \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) \vec{x}_k^T \vec{x}_{k'} + b, \end{aligned} \quad (23)$$

where $\vec{x}_{k'}$ is the new vector whose output is to be predicted, T represents the transpose operation on vectors and α_k and α_k^* are the Lagrange multipliers for the k th observation.

In most cases, there is a non-linear relationship between input and output variables and a non-linear support vector regression algorithm is needed. In order to make the support vector

model non-linear, the input vectors are mapped into a higher dimensional feature space using a mapping function, $\phi(x)$. However, in most of the cases, explicit mapping results in infeasible solutions that are computationally hard to obtain. The feasible way to convert a linear SMV's into non-linear SMV is to use kernel mapping which maps the input vectors into a higher dimensional feature space, i.e., $k(x, x') = \langle \phi(x), \phi(x') \rangle$ which changes the SMV algorithm as:

$$\begin{aligned} \max_{\alpha, \alpha^*} &= \frac{1}{2} \sum_{k, k'=1}^{nd} (\alpha_k - \alpha_k^*) (\alpha_{k'} - \alpha_{k'}^*) k(\vec{x}_k, \vec{x}_{k'}) \\ &\quad - \varepsilon \sum_{k=1}^{nd} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) y_k \\ \text{subject to} &\quad \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) = 0 \\ &\quad \alpha_k, \alpha_k^* \in [0, C]. \end{aligned} \quad (24)$$

Note that one may choose various different kernel functions, e.g., Gaussian radial base kernel, polynomial kernel, satisfying the Mercer's condition (Vapnik 1998) and the output value of the k' th input vector is calculated using the following function:

$$\begin{aligned} \hat{y}_{k'} &= \hat{f}(\vec{x}_{k'}, \alpha, \alpha^*) \\ &= \sum_{k=1}^{nd} (\alpha_k - \alpha_k^*) K(\vec{x}_{k'}, \vec{x}_k) + b, \end{aligned} \quad (25)$$

where the $K(\vec{x}_{k'}, \vec{x}_k)$ represents the kernel mapping of the input vectors. From Eq. (25), the dependent variable is estimated using the kernel mapping of the input vectors and their Lagrange multipliers of each vector that are calculated from the optimization algorithm. Note that in Eq. (25), it is not required to calculate the weight vector explicitly, the input vectors whose Lagrange multipliers are not zero are used in estimating the output and they are called the “**support vectors**”. In a sense, the complexity of the functions, i.e., the $\frac{1}{2} \|\vec{w}\|^2$ term, represented by support vectors is independent of the dimensionality of the input

space x , and only depends on the number of support vectors.

An additional result of the application of Karush–Kuhn–Tucker (KKT) theorem in SVR is:

$$\begin{aligned} (C - \alpha_k)\zeta_k &= 0 \\ (C - \alpha_k^*)\zeta_k^* &= 0. \end{aligned} \quad (26)$$

One of the several conclusions one might make from (26) is that $\alpha_k\alpha_k^* = 0$, i.e., there can never be a set of dual variables for an observation k which are both simultaneously non-zero as this would require non-zero slacks in both directions. Since $c > 0$, then $\zeta_k\zeta_k^* = 0$ must also be true. In the same sense, there can never be two slack variables $\zeta_k, \zeta_k^* > 0$ which are both non-zero and equal.

Special Fuzzy Functions with SVM (FF-SVM) Method

As one can build ordinary least squares for the estimation of the Special Fuzzy Functions when the relationship between input variables and the output variable can be linearly defined in the original input space, one may also build support vector regression models to estimate the parameters of the non-linear Special Fuzzy Functions using support vector regression methods. The augmented input matrix is determined from FCM algorithm such that there is one SVR in SFF-SVM for each cluster same as SFF-LSE model. One may choose any membership transformation depending on the input dataset. Then one can apply support vector regression, SVR, algorithm instead of LSE to each augmented matrix, which are comprised of the original selected input variables and the membership values and/or their transformations. Support vector machines' optimization algorithm is applied to each augmented matrix of each cluster (rule) i , $i = 1, \dots, c^*$, to optimize their Lagrange multipliers, α_{ik} and α_{ik}^* , and find the candidate support vectors, $k = 1, \dots, nd$. Hence, using SFF-SVM, one finds Lagrange multipliers of each k th train data sample one for each cluster, i . Then the output value of k th data sample in i th cluster is estimated using the Eq. (27) as follows:

$$\begin{aligned} \hat{y}_{ik'}^* &= \hat{f}(\vec{x}_{ik'}, \alpha_i, \alpha_i^*) \\ &= \sum_{k=1}^{nd} (\alpha_{ik} - \alpha_{ik}^*) K(\vec{x}_{ik'}, \vec{x}_{ik}) + b_i. \end{aligned} \quad (27)$$

Where the $\hat{y}_{ik'}^*$ is the estimated output of the k' th vector in i th cluster which is estimated using the support vector regression function with the Lagrange multipliers of the i th cluster. The augmented kernel matrix denotes the kernel mapping of the augmented input matrix (as described in SFF-LSE approach) where the membership values and their transformations are used as additional input variables. After the optimization algorithm finds the optimum Lagrange multipliers, one can estimate the output value of each data point in each cluster using Eq. (27).

The inference structure of SFF-SVM is adapted from the Special Fuzzy Functions with least squares where one can estimate a single output of a data point (see Eq. (15) by taking the membership value weighted averages of its output values calculated for each cluster using Eq. (27).

Future of Fuzzy System Models

In the future, fuzzy system models are expected to be structured by type 2 fuzzy sets. For this purpose, we next present basic definitions.

Basic Definitions

Definition 1 A type 2 fuzzy set $\tilde{A}$ on universe of discourse, x , is a fuzzy set which is characterized by a fuzzy membership function, $\tilde{\mu}_A(x)$, where $\tilde{\mu}_A(x)$ is a mapping as shown below:

$$\tilde{\mu}_A(x) : X \rightarrow [0, 1]^{[0,1]}. \quad (28)$$

Then type 2 fuzzy set, $\tilde{A}$, can be characterized as follows:

$$\tilde{A} = \{(x, (u, f_x(u))) | x \in X, u \in J_x \subseteq [0, 1]\}. \quad (29)$$

Where u is defined as the primary membership value, $J_x \subseteq [0, 1]$ is the domain of u and $f_x(u)$ is the secondary membership function. An alternative

definition of type 2 fuzzy set $\tilde{A}$, used by Mendel (2001) and inspired by Mizumoto and Tanaka (1976), can be given as follows:

Given that x is a continuous universe of discourse, the same type 2 fuzzy set can be defined as:

$$\tilde{A} = \int_{x \in X} \left[\int_{u \in J_x} f_x(u)/u \right] / x. \quad (30)$$

And for discrete case, the same type 2 fuzzy set can be defined as:

$$\tilde{A} = \sum_{x \in X} \left[\sum_{u \in J_x} f_x(u)/u \right] / x. \quad (31)$$

The secondary membership function, $f_x(u)$, can be defined as follows:

Definition 2 Secondary membership function, $f_x(u)$, is a function that maps membership values of universe of discourse x onto unit interval $[0, 1]$. Thus, $f_x(u)$ can be characterized as follows:

$$f_x(u) : J_x \rightarrow [0, 1], \forall u \in J_x \subseteq [0, 1], \quad \forall x \in X. \quad (32)$$

With the secondary membership function defined as above, the membership function of type 2 fuzzy set $\tilde{A}$, $\tilde{\mu}_A(x)$, can then be written for continuous and discrete cases, respectively, as follows:

$$\tilde{\mu}_A(x) = \int_{u \in J_x} f_x(u)/u, \quad \forall x \in X, \quad (33)$$

$$\tilde{\mu}_A(x) = \sum_{u \in J_x} f_x(u)/u, \quad \forall x \in X. \quad (34)$$

An Interval Valued Type 2 Fuzzy Set (IVT2FS), which is a special case of type 2 fuzzy set, can be defined as follows:

Definition 3 Let $\tilde{A}$ be a linguistic label with type-2 membership function on the universe of discourse of base variable x , $\tilde{\mu}_A(x) : X \rightarrow f_x(u)/u, u \in J_x, J_x \subseteq [0, 1]$. The following condition needs to be satisfied in order to consider $\tilde{\mu}_A(x)$ as an interval value type 2 membership functions:

$$f_x(u) = 1, \forall x \in X, \forall u \in J_x, J_x \subseteq [0, 1]. \quad (35)$$

Thus, the interval valued type 2 membership function is a mapping as shown below:

$$\tilde{\mu}_A(x) : X \rightarrow 1/u, u \in J_x, J_x \subseteq [0, 1]. \quad (36)$$

General Structure of Type 2 Fuzzy System Models

In a series of papers, Mendel, Karnik and Liang (Karnik and Mendel 1998a, b, 2000; Karnik et al. 1999) extended traditional type 1 inference methods such that these methods can process type 2 fuzzy sets. These studies were explained thoroughly by Mendel in (2001). The classical Zadeh and Takagi–Sugeno type 1 models are modified as type 2 fuzzy rule bases (T2Z-FR and T2TS-FR), respectively, as follows:

$$\begin{aligned} & \text{ALSO}_{i=1}^{c^*} \left[\text{IF AND}_{j=1}^{NV} \left(x_j \in X_j \text{ isr } \tilde{A}_{ji} \right) \right. \\ & \quad \left. \text{THEN } y \in Y \text{ isr } \tilde{B}_i \right]. \end{aligned} \quad (37)$$

$$\begin{aligned} & \text{ALSO}_{i=1}^{c^*} \left[\text{IF AND}_{j=1}^{NV} \left(x_j \in X_j \text{ isr } \tilde{A}_{ji} \right) \right. \\ & \quad \left. \text{THEN } y_i = a_i x^T + b_i \right]. \end{aligned} \quad (38)$$

Mendel, Karnik and Liang (Karnik and Mendel 1998a, b, 2000; Karnik et al. 1999) assumed that the antecedent variables are separable (i.e., non-interactive). After formulating the inference for a full type 2 fuzzy system model, Karnik et al. (Karnik and Mendel 1998a, b, 2000; Karnik et al. 1999) simplified their proposed methods for the interval values case. In order to identify the structure of the IVT2FS, it was assumed that the membership functions are Gaussian. A clustering method was utilized to identify the mean parameters for the Gaussian functions. However, the clustering method has not been specified. It was assumed that the standard error parameters of the Gaussian membership functions are exactly known. The number of

rules was assigned as eight due to the nature of their application. However, the problem of finding the suitable number of rules was not discussed in the paper.

Liang and Mendel (2000) proposed another method to identify the structure of IVT2FS. It has been suggested to initialize the inference parameters and to use steepest-descent (or other optimization) method in order to tune these parameters of an IVT2FS. Two different approaches for the initialization phase have been suggested in (Liang and Mendel 2000). Partially dependent approach utilizes a type 1 fuzzy system model to provide a baseline for the type 2 fuzzy system model design. Totally independent approach starts with assigning random values to initialize the inference parameters. Liang and Mendel (2000) indicated that the main challenge in their proposed tuning method is to determine the active branches.

Mendel (2001) indicated that several structure identification methods, such as one-pass, least-squares, back-propagation (steepest descent), singular value-QR decomposition, and iterative design methods, can be utilized in order to find the most suitable inference parameters of the type 2 fuzzy system models. Mendel (2001) provided an excellent summary of the pros and cons of each method.

Several other researchers such as Starczewski and Rutkowski (2002), John and Czarnecki (1998, 1999) and Chen and Kawase (2000) worked on T2-FSM. Starczewski and Rutkowski (2002) proposed a connectionist structure to implement interval valued type 2 fuzzy structure and inference. It was indicated that methods such as, back propagation, recursive least squares or Kalman algorithm-based methods, can be used in order to determine the inference parameters of the structure. John and Czarnecki (1998, 1999) extended the ANFIS structure such that it can process the type 2 fuzzy sets.

All of the above methods assume non-interactivity between antecedent variables. Thus, the general steps of the inference can be listed as: fuzzification, aggregation of the antecedents,

implication, and aggregation of the consequents, type reduction and defuzzification.

With the general structure of type 2 fuzzy system models and inference techniques, we next propose discrete interval valued type 2 rule base structures.

Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS)

In Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS) (Uncu and Türkşen 2007b) the domain of the primary membership is discrete and the secondary membership values are fixed to 1. Thus, the proposed method is utilizing discrete interval valued type 2 fuzzy sets in order to represent the linguistic values assigned to each fuzzy variable in each rule. These fuzzy sets can be mathematically defined as follows:

$$\tilde{A} = \int_{x \in X} \left[\sum_{u \in J_x} 1/u \right] / x, \quad (39)$$

where $x \in X \subseteq \mathfrak{R}$, $u \in J_x = \{J_{xr}\}$, $r = 1, \dots, NM$, x is a system variable with continuous domain, u is the primary membership value with discrete domain and J_{xr} is the r th primary membership value associated with x . The discrete interval valued type 2 fuzzy set $\tilde{A}$, can be considered as the union of type 1 fuzzy sets a^r , $r = 1, \dots, nm$, where nm is the number of embedded type 1 fuzzy sets that form the “discrete interval valued type 2 fuzzy set $\tilde{A}$ ”. (for the purposes of this paper, we assume that nm is known. Currently we are conducting further research to determine nm at the Knowledge-Intelligence Laboratory, University of Toronto). Consequently, the membership function of discrete interval type 2 fuzzy set $\tilde{A}$, $\mu_{\tilde{A}}(x)$, can be represented as a collection of type 1 membership functions as follows:

$$\mu_{\tilde{A}}(x) = \{\mu_A^r(x)\}, r = 1, \dots, NM, \quad (40)$$

where nm is the number of discrete membership values assigned to each value of system variable x , $\mu_A^r(x)$ is the membership function associated

with r th embedded type 1 fuzzy set, which can be defined as follows:

$$\mu_A^r(x) : X \rightarrow J_{x^r}, x \in X \subseteq \mathfrak{R}. \quad (41)$$

One of the goals of this study is to eliminate the non-interactivity assumption used in the existing type 2 fuzzy system models. Hence, we have extended the fuzzy rule base structure proposed by Delgado et al. (1997), Babuska and Verbruggen (1997) and Uncu and Türkşen (2007b). Thus, the proposed Discrete Interval Type 2 Zadeh Fuzzy Rule base (DIT2-Z-FR) structure can be written as follows:

$$R : \left\{ \text{ALSO}_{i=1}^{c^*} (\text{IF } x \in X \text{ isr } A_i^r \text{ THEN } y \in Y \text{ isr } B_i^r) \right\}, \\ r = 1, \dots, NM, \quad (42)$$

where, A_i^r is the r th embedded type 1 multi-dimensional fuzzy set associated with the antecedent of the i th rule. A_i^r will be represented with the membership function $\mu_i^r(x)$ in the membership value domain. Similarly, B_i^r is the r th embedded type 1 fuzzy set associated with the consequent of the i th rule. B_i^r will be represented with the membership function $\mu_i^r(y)$ in the membership value domain.

The rule base structure given in (42) can be thought as a collection of embedded type 1 fuzzy rule bases. Hence, the proposed fuzzy rule base structure is named as Discrete Interval Type 2 Fuzzy Rule base (DIT2-Z-FR) structure.

The other well-known fuzzy rule base structures are also extended to type 2 by using the above idea. Hence, Discrete Interval Valued Type 2 Takagi–Sugeno Fuzzy Rule base (DIT2-TS-FR) can be written respectively as follows:

$$R : \left\{ \text{ALSO}_{i=1}^{c^*} (\text{IF } x \in X \text{ isr } A_i^r \text{ THEN } y = a_i^r x^T + b_i^r) \right\}, \\ r = 1, \dots, NM, \quad (43)$$

where, $a_i^r = (a_{i,1}^r, \dots, a_{i,NV}^r)$ is the regression coefficient vector associated with the i th rule of the r th embedded type 1 fuzzy system model and b_i^r is the scalar associated with the i th rule of the r th embedded type 1 fuzzy system model. As one can observe the Takagi–Sugeno structure is not only extended by representing the antecedent fuzzy sets with discrete type 2 fuzzy sets but also by letting uncertainty in crisp inference parameters in consequents. Thus, the problem of building type 2 fuzzy system models is reduced to finding embedded type 1 fuzzy system models.

Discrete Interval Valued Type 2 “Special Fuzzy Functions”, (SFF)

In a similar manner to Discrete Interval Valued Type 2 Fuzzy Rule bases, we could construct Discrete Interval Valued Type 2 “Special Fuzzy Functions” for the cases of SFF-LSE and SFF-SVR. For example, the function of a single input single output model, which includes only the membership values as the additional input variable, $Y_i^r = \beta_{i0}^r + \beta_{i1}^r \Gamma_i^r + \beta_{i2}^r X$, $r = 1, \dots, NM$ that represents the r th “Special Fuzzy Function” to the i th interactive (joint) type 2 cluster in (Y_i^r, Γ_i^r, X^r) , $i = 1, \dots, c^*$, $r = 1, \dots, NM$ space, would be estimated with SFF-LSE approach as follows:

$$\beta_i^{r*} = (X_i^{rT} X_i^{r'})^{-1} (X_i^{rT} Y_i^r), r = 1, \dots, NM, \quad (44)$$

where $\beta_i^{r*} = (\beta_{i0}^{r*}, \beta_{i1}^{r*}, \beta_{i2}^{r*})$ and $X_i^{r'} = [1, \Gamma_i^r, X]$, provided the inverse of covariance, $(X_i^{rT} X_i^{r'})^{-1}$, exists. The estimate of y_i would be obtained as:

$$Y_i^{r*} = \beta_{i0}^{r*} + \beta_{i1}^{r*} \Gamma_i^r + \beta_{i2}^{r*} X. \quad (45)$$

The single output value is calculated using each output value, one from each cluster, and weighting them with their corresponding membership values as follows:

$$Y_i^{r*} = \frac{\sum_{i=1}^{c^*} \gamma_i^r Y_i^{r*}}{\sum_{i=1}^{c^*} \gamma_i^r}, i = 1, \dots, c^*, r = 1, \dots, NM. \quad (46)$$

The development of such Discrete Interval Valued Type 2 “Special Fuzzy Functions” with LSE will be studied in our future investigations. In a similar manner, we propose to develop Discrete Interval Valued Type 2 “Special Fuzzy Functions” with SVM methodology.

Case Study Applications

In order to test the proposed model performances as opposed to fuzzy rule base systems three input-output datasets are considered in this investigation. These are:

- (i) Daily price of a stock in stock market.
- (ii) Customer Income Prediction model for a major bank.
- (iii) The amount of chemicals for a desulphurization process for a steel processing company.

The specifications of each datasets are displayed in the following parts:

Daily Stock Price Dataset

Daily stock price dataset comprises of the daily trend data of stock prices. This dataset was introduced by Sugeno and Yasukawa (1993). The same dataset has been used in various other studies one of which compares six different fuzzy reasoning methods using this dataset.

Out of 100 observations, 50 of them are used for the training purposes and the other 50 was hold-out for testing purposes. There were originally 11 input variables and single output variable in the dataset (Uncu and Türkşen 2007b). Preliminary input selection was applied using Random Forests (RF) method, (Uncu and Türkşen 2007b) which estimates variable importance. Based on the results of RF method, only 4 of input variables

i.e., x_2, x_4, x_8, x_{10} , were found to have importance on the output variable. The rest of the variables had insignificant effect on the output.

Income Prediction Dataset

The purpose of the Income Prediction Model was to predict the income of the future customers based on the current customer information and 1996 year census data. There were more than 200 variables and hundred of thousands of customers. According to business needs, the dataset was partitioned into 9 different parts based on age, number of different types of investment accounts hold by the customer and different regions of residency. In this study, only one partition was investigated.

We have only selected 10% of the i.i.d. data samples to do our research on using only single partition explained above. The data was cleaned from the outliers using the expert’s knowledge and 900 training and 900 testing observations are selected randomly. The dataset was comprised of 11 input variables (Uncu and Türkşen 2007b). Based on the correlation analysis, 3 input variables were discarded from the dataset resulting in 8 input variables. There were no census variables among the selected input variables.

Desulphurization Dataset

A torpedo car desulphurization facility removes sulfur from hot metal leaving the blast furnaces before it is sent to the next process. Generally, desulphurization is carried out by injecting two different powered reagents directly into the hot metal via a lance. The reagents react with the sulfur in the hot metal and residue, which is rich in sulfur, is separated from the iron.

The aim of the data-mining project was to determine the right amounts of the reagents to be added into the hot metal. These reagents are expensive materials and precise estimation is required. There are vast quantities of data available on the desulphurization process, which has various characteristics. The original input and output variables are shown in (Mamdani and Assilian 1981). There are 750 training and 900 verification samples used in the experiments. Based in the variable selection

using random forest regression method, only 5 variables are found to be important.

Experimental Design

Using the three different input-output datasets, we build four different fuzzy system model structures, SFF-LSE, SFF-SVM, SY-FRB and TS-FRB. In order to keep the consistency between each model structure, the same training and testing datasets are used for the four fuzzy system models with the same input variables. The categorical variables are transformed into probabilities using logistic regression and are used as additional inputs only in Income Prediction and Desulphurization Datasets in all of the four models.

Sugeno–Yasukawa Models and Takagi–Sugeno Models

The proposed FSM models are compared to two well known Fuzzy Rule Base Models: (i) Sugeno and Yasukawa’s fuzzy logic based approach using Partition Type Fuzzy Model, **SY-FRB**, (Sugeno and Yasukawa 1993) (ii) Takagi and Sugeno’s fuzzy system modeling approach, **TS-FRB** (Takagi and Sugeno 1985). In Sugeno and Yasukawa’s FSM approach, they use linguistic variables for both the consequent and the antecedent part of the fuzzy rules and the system learns all inference parameters from the data without the expert intervention. The variable selection method defined in their paper is not applied to these 3 datasets in order to compare the models on the same basis.

In Takagi and Sugeno’s FRB (TS-FRB) structure (Takagi and Sugeno 1985), they assume that the antecedent membership functions are to be characterized with triangular membership functions. In their approach, each input variable space is assumed to be partitioned into two clusters and logical connective AND is taken as MIN. Then, the structure identification problem is just to identify the regression equation coefficients for each rule and the antecedent parameters for each input variable in each rule. Researchers proposed several structure identification methods to identify the membership functions from the data, e.g.,

Delgado et al. (1997), Babuska and Verbruggen, etc. In this paper we have used Babuska et al.’s (Babuska and Verbruggen 1997) modified Takagi–Sugeno study where the membership functions of the antecedents are identified using fuzzy c-means clustering and projected onto each input vector. The degree of fulfillment of each rule is then calculated using a t-norm operator, i.e., product. Only one aggregate input membership function is identified for each rule. The inference parameters are same as the traditional Takagi–Sugeno inference method (Takagi and Sugeno 1985).

Special Fuzzy Functions Models (SFFM)

In this paper, the modeling performance from the Special Fuzzy Functions with LSE and SVM models, SFF-LSE and SFF-SVM, are compared to the Fuzzy Rule Base structures. Instead of using cluster validity indices to select the optimum model parameters, we measured the optimum model based on the best model performance using RMSE by applying a grid search for each parameter. Note that fuzzy functions with LSE system models have 2 parameters, which are the FCM parameters, i.e., degree of fuzziness and cluster size. Special Fuzzy Functions with SVR models have 4 parameters: FCM parameters and the SVR parameters which are *C-regularization* and *ϵ -insensitive value (epsilon)*.

In both of these special fuzzy function models, membership values and their exponential transformations are used as additional input variables. We applied the LIBSVM program (Chang and Lin 2001) within our special fuzzy function codes for support vector optimization in estimating the “Special Fuzzy Functions”. We chose Gaussian RBF as the kernel function, $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ in all the experiments and the default values of the kernel parameters in LIBSVM (Chen and Kawase 2000) are used. Recall that the SVR regression has two parameters that is set by the user: ϵ -insensitive zone (*epsilon* or ϵ) and the regularization parameter, C . The parameters of SVM regression, C and ϵ , are generated by the grid search, i.e., $C = \{2^{-3}, 2^{-1}, \dots, 2^3, 2^5\}$, $\epsilon = \{0.1, \dots, 0.5\}$, as well as the FCM parameters, i.e., $c =$

3, 5, ..., 10, $m = 1.1, \dots, 2.5$ in all the experiments. Hence, for the fuzzy functions with LSE models 2 parameters are specified for each model, and for the fuzzy functions using SVM models, 4 parameters, m , c , C , and γ are determined using a grid search where the model performance of each model is determined using Root Mean Square Error of the models as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (47)$$

y_i , and $\hat{y}_i$ are the actual and estimated output values of a single observation, N is the total number of observations in the dataset.

Experimental Model Results

The 4 fuzzy system models are applied on the daily stock price of a stock market, income prediction, and reagents estimation in desulphurization process datasets. The results are displayed in Tables 1, 2 and 3.

Special Fuzzy Function (SFF) models, when estimated with either *SVM* or *LSE* algorithms, show better generalization than the fuzzy rule base models. The fuzzy function models can

increase the model performance by up to 35% depending on the dataset.

Table 4 displays the optimum parameters of the models from each experiment whose results are displayed in Tables 1, 2 and 3. In Table 4, $C\text{-reg}$ indicates the regularization parameter, ε is the ε -insensitive region (epsilon), m refers to the degree of fuzziness (weighting exponent) of the fuzzy c -mean clustering algorithm, c indicates the number of cluster, and $\#sv$ refers to the average number of support vectors from each support vector regression model build for each cluster of the *SFF-SVM* models. In order to show the dispersion of the number of support vectors among each cluster we also included the standard deviation (σ_{sv}) of the support vectors of the optimum models.

The grid search algorithms applied in this paper try to find the best RMSE value from training data in each experiment and assign these parameters as the optimum model parameters. The algorithm searches for the minimum regression error. Then, verification dataset output is inferred using the optimum parameters. The issue with these grid search algorithms is that, sometimes, the models get stuck in the local minimum which is smaller than the global minimum and this might cause generalization problems. An example to this concept is shown in income prediction dataset (Table 2). The model parameters best fit to the training data when FF-SVM is used but this causes generalization problems. It should also be reminded that, when there is a

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 1 Daily stock price of a stock in stock market

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
RMSE (train)	2.43	3.82	2.76	7.16
RMSE (test)	3.64	5.61	5.68	9.93

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 2 Income prediction dataset*

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
RMSE (train)	0.40	0.52	0.49	0.58
RMSE (test)	0.64	0.64	0.80	0.70

* The RMSE values are calculated from standardized output values

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 3 Reagent estimation for desulphurization process

	SFF-SVM	SFF-LSE	TS-FRB	SY-FRB
Reagent1				
RMSE (train)	30	40	35	69
RMSE (test)	42	45	45	72
Reagent2				
RMSE (train)	4.80	6.49	5.62	10.01
RMSE (test)	6.59	7.19	7.19	10.80

Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions, Table 4 Optimum model parameters of three datasets

Dataset	Model type	Optimum model parameters
Daily stock price	SFF-SVM	$C - reg = 32, \varepsilon = 0.2, c = 8, m = 1.9, \# sv = 28(\sigma_{sv} = 1.5)$
	SFF-LSE	$c = 8, m = 1.6$
Income prediction	SFF-SVM	$C - reg = 64, \varepsilon = 0.2, c = 7, m = 1.2, \# sv = 264(\sigma_{sv} = 5.7)$
	SFF-LSE	$c = 8, m = 1.6$
Desulphurization	SFF-SVM	Reagent 1: $C - reg = 64, \varepsilon = 0.2, c = 7, m = 1.4, \# sv = 121(\sigma_{sv} = 4.8)$ Reagent 2: $C - reg = 64, \varepsilon = 0.2, c = 7, m = 1.5, \# sv = 101(\sigma_{sv} = 7.7)$
	SFF-LSE	Reagent 1: $c = 5, m = 1.4$ Reagent 2: $c = 6, m = 1.5$

linear relationship between the inputs and the output, then LSE model performances will be as good as the other model performances. On the other hand, FF-LSE models, in three of the datasets, show more reliable results than the SVM models. One should run both models and determine the optimum model parameters after observing the results from both models.

Conclusions and Future Directions

In this paper, we have outlined basic well known three type 1 system models of the recent past and suggested that type 2 fuzzy system models are likely to be studied more extensively in the future. In particular, we have reviewed: (1) Type 1 Fuzzy Rule bases and (2) Type 1 “Special Fuzzy Functions”. Furthermore we discussed the Type 2 Fuzzy Rule bases. But we left the Type 2 “Special Fuzzy Functions” for a future study after providing a structure for their development. As well, we have demonstrated that “Type 1 Special Fuzzy Functions” provide better results than “Type 1 Fuzzy Rule Base” models in three specific case studies.

Bibliography

Primary Literature

Babuska R, Verbruggen HB (1997) Constructing fuzzy models by product space clustering. In: Hellendoorn H, Driankov D (eds) Fuzzy model identification: selected approaches. Springer, Berlin, pp 53–90

- Bezdek JC (1973) Fuzzy mathematics in pattern classification. Ph.D thesis, Applied Mathematics Center. Cornell University, Ithaca
- Celikyilmaz A, Türkşen IB (2007) Fuzzy functions with support vector machines. Inf Sci 177:5163–5177
- Celmins A (1987a) Least squares model fitting to fuzzy vector data. Fuzzy Sets Syst 22:245–269
- Celmins A (1987b) Multidimensional least squares model fitting of fuzzy models. Math Model 9:669–690
- Chang YHO, Ayyub BM (1993) Reliability analysis in fuzzy regression. In: Proceedings of annual conference of NAFIPS’93. IEEE, Allentown/New York, pp 93–97
- Chang PT, Lee ES (1994) Fuzzy linear regression with spreads unrestricted in sign. Compt Math Appl 28: 61–71
- Chang C, Lin C (2001) LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/%7Ecjlin/libsvm>
- Chen Q, Kawase S (2000) On fuzzy-valued fuzzy reasoning. Fuzzy Sets Syst 113:237–251
- Chen MS, Wang SW (1999) Fuzzy clustering analysis for optimizing fuzzy membership functions. Fuzzy Sets Syst 103(2):239–254
- Chiu S (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2(3):267–278
- Delgado M, Gomez-Skermata AF, Martin F (1997) Rapid prototyping of fuzzy models. In: Hellendoorn H, Driankov D (eds) Fuzzy model identification: selected approaches. Springer, Berlin, Germany, pp 53–90
- Demirci M (1999) Fuzzy functions and their fundamental properties. Fuzzy Sets Syst 106:239–246
- Demirci M (2003) Foundations of fuzzy functions and vague algebra based on many-valued equivalence relations, part I: fuzzy functions and their applications. IJ Gen Syst 32:123–155
- Demirci M, Recasens J (2004) Fuzzy groups, fuzzy functions and fuzzy equivalence relations. Fuzzy Sets Syst 144:441–458
- Diamond P (1998) Fuzzy least squares. Inf Sci 46:141–157
- Emami MR, Türkşen IB, Goldenberg AA (1998) Development of a systematic methodology of fuzzy logic modeling. IEEE Tran Fuzzy Syst 6(3):346–361
- Hathaway RJ, Bezdek JC (1993) Switching regression models and fuzzy clustering. IEEE Trans Fuzzy Syst 1(3):195–203

- Jang JSR (1993) Anfis: adaptive-network-based fuzzy inference systems. *IEEE Trans Syst Man Cybern* 23(3):665–685
- John RI, Czarnecki C (1998) A type 2 adaptive fuzzy inference system. In: *Proceedings of IEEE conference systems, man and cybernetics*, vol 2. IEEE, New York, pp 2068–2073
- John RI, Czarnecki C (1999) An adaptive type-2 fuzzy system for learning linguistic membership grades. In: *Proceedings of IEEE international fuzzy systems conference*, vol 3. IEEE, New York, pp 1552–1556
- Karnik NN, Mendel JM (1998a) Introduction to type-2 fuzzy logic systems. In: *Proceedings of IEEE conference on computational intelligence*, vol 2. IEEE, New York, pp 915–920
- Karnik NN, Mendel JM (1998b) Type-2 fuzzy logic systems: type reduction. In: *Proceedings of IEEE conference on systems, man and cybernetics*, vol 2. IEEE, New York, pp 2046–2051
- Karnik NN, Mendel JM (2000) Applications of type-2 fuzzy logic systems: handling the uncertainty associated with surveys. In: *Proceedings of IEEE conference on fuzzy systems*, vol 3, pp 1546–1551
- Karnik NN, Mendel JM, Liang Q (1999) Type-2 fuzzy logic systems. *IEEE Trans On Fuzzy Syst* 7(6): 643–658
- Liang Q, Mendel JM (2000) Interval type-2 fuzzy logic systems: theory and design. *IEEE Trans On Fuzzy Syst* 8(5):535–550
- Mamdani EH, Assilian S (1981) An experiment in linguistic synthesis with a fuzzy logic controller. In: Mamdani EH, Gains BR (eds) *Fuzzy reasoning and its applications*. Academic Press, New York
- Mendel JM (2001) *Uncertain rule-based fuzzy logic systems: introduction and new directions*. Prentice, Upper Saddle River
- Mizumoto M (1989) Method of fuzzy inference suitable for fuzzy control. *J Soc Instrum Control Eng* 58: 959–963
- Mizumoto M, Tanaka K (1976) Some properties of fuzzy sets of type 2. *Inf Control* 31:312–340
- Nakanishi H, Türkşen IB, Sugeno M (1993) A review and comparison of six reasoning methods. *Fuzzy Sets Syst* 57:257–295
- Pal NR, Bezdek JC (1995) On cluster validity for the fuzzy c-means model. *IEEE Trans Fuzzy Syst* 3(3):370–379
- Rutkowska D (2002) Type 2 fuzzy neural networks: an interpretation based on fuzzy inference neural networks with fuzzy parameters. In: *Proceedings of IEEE conference on fuzzy systems*, vol 2. IEEE, New York, pp 1180–1185
- Savic D, Pedrycz W (1991) Evolution of fuzzy linear regression models. *Fuzzy Sets Syst* 39:51–63
- Smola AJ, Scholkopf B (1998) A tutorial on support vector regression. *NeuroCOLT2 technical report series*, NC2-Tr-1998–030
- Sproule BA, Bazoon M, Shulman KI, Türkşen IB, Naranjo CA (2000) Fuzzy logic pharmacokinetic modeling: an application to lithium concentration prediction. *Clin Pharmacol Ther* 62:29–40
- Starczewski J, Rutkowski L (2002) Connectionist structures of type 2 fuzzy inference systems. In: Wyrzykowski R et al (eds) *PPAM 2001*, LCNS 2328. Springer, Heidelberg, pp 634–642
- Sugeno M, Yasukawa T (1993) A fuzzy logic based approach to qualitative modeling. *IEEE Trans Fuzzy Syst* 1:7–31
- Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans Syst Man Cybern SMC-15*(1):116–132
- Tanaka H, Vegima S, Asai K (1982) Linear regression analysis with fuzzy model. *IEEE Trans Syst Man Cybern SMC-2*:903–907
- Tanaka H, Ishibuchi H, Yoshikawa S (1995) Exponential possibility regression analysis. *Fuzzy Sets Syst* 69: 305–318
- Türkşen IB (1986) Interval valued fuzzy sets based on normal forms. *Fuzzy Sets Syst* 20:191–210
- Türkşen IB (1992) Interval-valued fuzzy sets and ‘compensatory AND’. *Fuzzy Sets Syst* 51:295–307
- Türkşen IB (1995) Fuzzy normal forms. *Fuzzy Sets Syst* 69:319–346
- Türkşen IB (2002) Type 2 representation and reasoning for CWW. *Fuzzy Sets Syst* 127:17–36
- Türkşen IB (2008) Fuzzy functions with LSE. *Appl Soft Comput* 8(3):1178–1182
- Uncu Ö, Türkşen IB (2007a) A novel feature selection approach: combining feature wrappers and filters. *Inf Sci* 177:449–466
- Uncu Ö, Türkşen IB (2007b) Discrete interval type 2 fuzzy system models using uncertainty in learning parameters. *IEEE Fuzzy Syst* 15(1):90–106
- Vapnik NV (1998) *Statistical learning theory*. Wiley, New York
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. *Inf Sci* 8: 199–249
- Zimmermann HJ, Zysno P (1980) Latent connectives in human decision-making. *Fuzzy Sets Syst* 4:37–51

Books and Reviews

- Kilic K (2002) A proposed fuzzy system modeling algorithm with an application in pharmacokinetic modeling. Ph.D thesis, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto

On Genetic-Fuzzy Data-Mining Techniques

Tzung-Pei Hong¹, Chun-Hao Chen² and Vincent S. Tseng³

¹Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, Taiwan

²Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan

³Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Data Mining

Fuzzy Sets

Fuzzy Data Mining

Genetic Algorithms

Genetic-Fuzzy Data-Mining Techniques

Multiobjective Genetic-Fuzzy Data Mining

Future Directions

Bibliography

Glossary

Data Mining: Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes. The common techniques include mining association rules, mining sequential patterns, clustering, and classification, among others.

Fuzzy Set Theory: The fuzzy set theory was first proposed by Zadeh in 1965. It is primarily concerned with quantifying and reasoning using natural language in which words can

have ambiguous meanings. It is widely used in a variety of fields because of its simplicity and similarity to human reasoning.

Fuzzy Data Mining: The concept of fuzzy sets can be used in data mining to handle quantitative or linguistic data. Basically, fuzzy data mining first uses membership functions to transform each quantitative value into a fuzzy set in linguistic terms and then uses a fuzzy mining process to find fuzzy association rules.

Genetic Algorithms: Genetic Algorithms (GAs) were first proposed by Holland in 1975. They have become increasingly important for researchers in solving difficult problems since they could provide feasible solutions in a limited amount of time. Each possible solution is encoded as a chromosome (individual) in a population. According to the principle of survival of the fittest, GAs generate the next population by several genetic operations such as crossover, mutation, and reproductions.

Genetic-Fuzzy Data Mining: Genetic algorithms have been widely used for solving optimization problems. If the fuzzy mining problem can be converted into an optimization problem, then the GA techniques can easily be adopted to solve it. They are thus called genetic-fuzzy data-mining techniques. They are usually used to automatically mine both appropriate membership functions and fuzzy association rules from a set of transaction data.

Multiobjective Genetic-Fuzzy Data Mining:

There are usually several criteria to be considered in solving a real problem. The multiobjective genetic algorithms are thus proposed for dealing with this issue. Thus, using multiobjective GA techniques for dealing with fuzzy data-mining problems is called multiobjective genetic-fuzzy data mining. They are usually used to automatically mine a set of Pareto solutions each of which represents an appropriate set of membership functions. The fuzzy association rules are then derived from a set of transaction data and the chosen membership functions.

Definition of the Subject and Its Importance

Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes. Most conventional data-mining algorithms identify the relationships among transactions using binary values. However, transactions with quantitative values are commonly seen in real-world applications. Fuzzy data-mining algorithms are thus proposed for extracting interesting linguistic knowledge from transactions stored as quantitative values. They usually integrate fuzzy-set concepts and mining algorithms to find interesting fuzzy knowledge from a given transaction data set. Most of them mine fuzzy knowledge under the assumption that a set of membership functions (Chen et al. 2007b) is known in advance for the problem to be solved. The given membership functions may, however, have a critical influence on the final mining results. Different membership functions may infer different knowledge. Automatically deriving an appropriate set of membership functions for a fuzzy mining problem is thus very important. There are at least two reasons for it. The first one is that a set of appropriate membership functions may not be defined by experts because lots of money and time are needed and experts are not always available. The second one is that data and concepts are always changing along with time. Some mechanisms are thus needed to automatically adapt the membership functions to the changes if needed. The fuzzy mining problem can thus be extended to finding both appropriate membership functions and fuzzy association rules from a set of transaction data.

Recently, genetic algorithms have been widely used for solving optimization problems. If the fuzzy mining problem can be converted into an optimization problem, then the GA techniques can easily be adopted to solve it. They are thus called genetic-fuzzy data-mining techniques. They are usually used to automatically mine both appropriate membership functions and fuzzy association rules from a set of transaction data. Some existing approaches are introduced

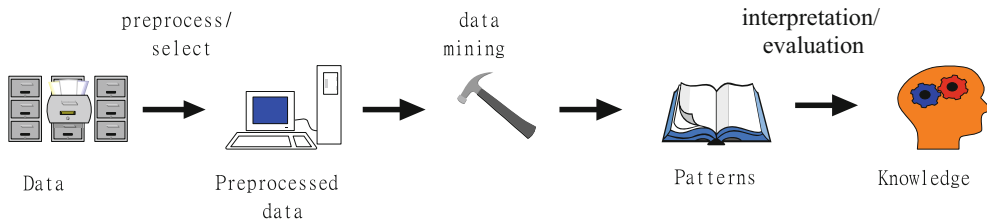
here. These techniques can dynamically adapt membership functions by genetic algorithms according to some criteria, use them to fuzzify the quantitative transactions, and find fuzzy association rules by fuzzy mining approaches.

Introduction

Most enterprises have databases that contain a wealth of potentially accessible information. The unlimited growth of data, however, inevitably leads to a situation in which accessing desired information from a database becomes difficult. Knowledge discovery in databases (*KDD*) has thus become a process of considerable interest in recent years, as the amounts of data in many databases have grown tremendously large. *KDD* means the application of nontrivial procedures for identifying effective, coherent, potentially useful, and previously unknown patterns in large databases (Frawley et al. 1991). The *KDD* process (Mamdani 1974) is shown in Fig. 1.

In Fig. 1, data are first collected from a single or multiple sources. These data are then pre-processed, such as sampling, feature selection or reduction, and data transformation, among others. After that, data-mining techniques are then used to find useful patterns, which are then interpreted and evaluated to form human knowledge.

Especially, data mining plays a critical role to the *KDD* process. It involves applying specific algorithms for extracting patterns or rules from data sets in a particular representation. Due to its importance, many researchers in the database and machine learning fields are primarily interested in this topic because it offers opportunities to discover useful information and important relevant patterns in large databases, thus helping decision-makers easily analyze the data and make good decisions regarding the domains concerned. For example, there may exist some implicitly useful knowledge in a large database containing millions of records of customers' purchase orders over the recent years. The knowledge can be found out by appropriate data-mining approaches. Questions such as "what are the most important trends in



On Genetic-Fuzzy Data-Mining Techniques, Fig. 1 A KDD process

customers' purchase behavior?" can thus be easily answered.

Most of the mining approaches were proposed for binary transaction data. However, in real applications, quantitative data exists and should also be considered. Fuzzy set theory is being used more and more frequently in intelligent systems because of its simplicity and similarity to human reasoning (Zadeh 1965). The theory has been applied in fields such as manufacturing, engineering, diagnosis, and economics, among others (Siler and James 2004; Zhang and Liu 2006). Several fuzzy learning algorithms for inducing rules from given sets of data have been designed and used to good effect with specific domains (Casillas et al. 2005; Hong and Lee 2001). As to fuzzy data mining, many algorithms are also proposed (Chan and Au 1997; Hong et al. 1999; Kuok et al. 1998; Lee et al. 2004; Yue et al. 2000).

Most of these fuzzy data-mining algorithms assume the membership functions are already known. In fuzzy mining problems, the given membership functions may, however, have a critical influence on the final mining results. Developing effective and efficient approaches to derive both the appropriate membership functions and fuzzy association rules automatically is thus worth to be studied. Genetic algorithms are widely used for finding membership functions in different fuzzy applications. In this entry, we discuss the genetic-fuzzy data-mining approaches which can mine both appropriate membership functions and fuzzy association rules (Chen et al. 2007a, b, 2008, 2009; Hong et al. 2004, 2006, 2008; Kaya and Alhadj 2005, 2006; Kaya 2006). The genetic-fuzzy mining problems can be divided into four kinds according to the types of fuzzy mining

problems and the ways of processing items. The types of fuzzy mining problems include Single-minimum-Support Fuzzy-Mining (SSFM) and Multiple-minimum-Support Fuzzy-Mining (MSFM). The ways of processing items include processing all the items together (integrated approach) and processing them individually (divide-and-conquer approach). Each of them will be described in detail in the following sections. But first of all, the basic concepts of data mining will be described below.

Data Mining

Data-mining techniques have been used in different fields to discover interesting information from databases in recent years. Depending on the type of databases processed, mining approaches may be classified as working on transaction databases, temporal databases, relational databases, multimedia databases, and data stream, among others. On the other hand, depending on the classes of knowledge derived, mining approaches may be classified as finding association rules, classification rules, clustering rules, and sequential patterns, among others.

Finding association rules in transaction databases is most commonly seen in data mining. It is initially applied to market basket analysis for getting relationships of purchased items. An association rule can be expressed as the form $A \rightarrow B$, where A and B are sets of items, such that the presence of A in a transaction will imply the presence of B . Two measures, support and confidence, are evaluated to determine whether a rule should be kept. The support of a rule is the fraction

of the transactions that contain all the items in A and B . The confidence of a rule is the conditional probability of the occurrences of items in A and B over the occurrences of items in A . The support and the confidence of an interesting rule must be larger than or equal to a user-specified minimum support and a minimum confidence, respectively.

To achieve this purpose, Agrawal and his coworkers proposed several mining algorithms based on the concept of large itemsets to find association rules in transaction data (Agrawal et al. 1993a, b, 1997; Agrawal and Srikant 1994). They divided the mining process into two phases. In the first phase, candidate itemsets were generated and counted by scanning the transaction data. If the number of an itemset appearing in the transactions was larger than a predefined threshold value (called minimum support), the itemset was considered a large itemset. Itemsets containing only one item were processed first. Large itemsets containing only single items were then combined to form candidate itemsets containing two items. This process was repeated until all large itemsets had been found. In the second phase, association rules were induced from the large itemsets found in the first phase. All possible association combinations for each large itemset were formed, and those with calculated confidence values larger than a predefined threshold (called minimum confidence) were output as association rules. In addition to the above approach, there are many other ones proposed for finding association rules.

Most mining approaches focus on binary valued transaction data. Transaction data in real-world applications, however, usually consist of quantitative values. Many sophisticated data-mining approaches have thus been proposed to deal with various types of data (Agrawal et al. 1993b; Srikant and Agrawal 1996; Zhang et al. 1997). It also presents a challenge to workers in this research field.

In addition to proposing methods for mining association rules from transactions of binary values, Srikant et al. also proposed a method (Srikant and Agrawal 1996) for mining

association rules from those with quantitative attributes. Their method first determines the number of partitions for each quantitative attribute and then maps all possible values of each attribute into a set of consecutive integers. It then finds large itemsets whose support values are greater than the user-specified minimum-support levels. For example, the following is a quantitative association rule “If *Age* is [20, . . . 29], then *Number of Car* is [0, 1]” with a support value (60%) and a confidence value (66.6%). It means that if the age of a person is from 20 to 29 years old, then they have zero or one car with 66.6%. Of course, different partition approaches for discretizing the quantitative values may influence the final quantitative association rules. Some researches have thus been proposed for discussing and solving this problem (Agrawal et al. 1993b; Zhang et al. 1997). Recently, fuzzy sets have also been used in data mining to handle quantitative data due to its ability to deal with the interval boundary problem. The theory of fuzzy sets will be introduced below.

Fuzzy Sets

Fuzzy set theory was first proposed by Zadeh in 1965 (Zadeh 1965). It is primarily concerned with quantifying and reasoning using natural language in which words can have ambiguous meanings. It is widely used in a variety of fields because of its simplicity and similarity to human reasoning (Chen et al. 2000; Siler and James 2004; Zhang and Liu 2006). For example, the theory has been applied in fields such as manufacturing, engineering, diagnosis, and economics, among others (Heng et al. 2006; Ishibuchi and Yamamoto 2005; Liang et al. 2002).

Fuzzy set theory can be thought of as an extension of traditional crisp sets, in which each element must either be in or not in a set. Formally, the process by which individuals from a universal set X are determined to be either members or non-members of a crisp set can be defined by a *characteristic or discrimination function* (Zadeh 1965). For a given crisp set A , this function assigns a value $\mu_A(x)$ to every $x \in X$ such that:

$$\mu_A(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{if and only if } x \notin A. \end{cases}$$

The function thus maps elements of the universal set to the set containing 0 and 1. This kind of functions can be generalized such that the values assigned to the elements of the universal set fall within specified ranges, referred to as the membership grades of these elements in the set. Larger values denote higher degrees of set membership. Such a function is called a membership function, $\mu_A(x)$, by which a fuzzy set A is usually defined. This function is represented by:

$$\mu_A : X \rightarrow [0, 1],$$

where $[0, 1]$ denotes the interval of real numbers from 0 to 1, inclusive. The function can also be generalized to any real interval instead of $[0, 1]$.

A special notation is often used in the literature to represent fuzzy sets. Assume that x_1 to x_n are the elements in fuzzy set A , and μ_1 to μ_n are, respectively, their grades of membership in A . A is then represented as follows:

$$A = \mu_1/x_1 + \mu_2/x_2 + \dots + \mu_n/x_n.$$

An α -cut of a fuzzy set A is a crisp set A_α that contains all the elements in the universal set X with their membership grades in A greater than or equal to a specified value of α . This definition can be written as:

$$A_\alpha = \{x \in X | \mu_A(x) \geq \alpha\}.$$

The scalar cardinality of a fuzzy set A defined on a finite universal set X is the summation of the membership grades of all the elements of X in A . Thus,

$$|A| = \sum_{x \in X} \mu_A(x).$$

Three basic and commonly used operations on fuzzy sets are complementation, union, and intersection, as proposed by Zadeh. They are described as follows.

1. The complementation of a fuzzy set A is denoted by $\neg A$, and the membership function of $\neg A$ is given by:

$$\mu_{\neg A}(x) = 1 - \mu_A(x), \quad \forall x \in X.$$

2. The intersection of two fuzzy sets A and B is denoted by $A \cap B$, and the membership function of $A \cap B$ is given by:

$$\mu_{A \cap B}(x) = \min \{\mu_A(x), \mu_B(x)\}, \quad \forall x \in X.$$

3. The union of two fuzzy sets A and B is denoted by $A \cup B$, and the membership function of $A \cup B$ is given by:

$$\mu_{A \cup B}(x) = \max \{\mu_A(x), \mu_B(x)\}, \quad \forall x \in X.$$

Note that there are other calculation formula for the complementation, union, and intersection, but the above are the most popular.

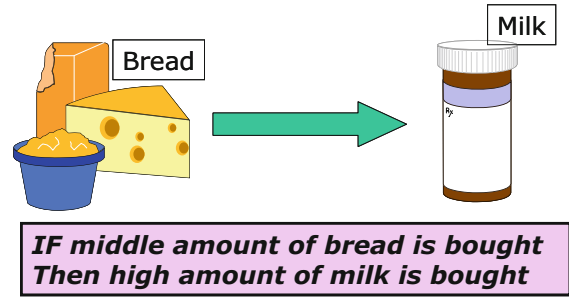
Fuzzy Data Mining

As mentioned above, the fuzzy set theory is a natural way to process quantitative data. Several fuzzy learning algorithms for inducing rules from given sets of data have thus been designed and used to good effect with specific domains (Casillas et al. 2005; Hong and Lee 2001; Rasmani and Shen 2004). Fuzzy data-mining approaches have also been developed to find knowledge with linguistic terms from quantitative transaction data. The knowledge obtained is expected easy to understand. A fuzzy association rules is shown in Fig. 2.

In Fig. 2, instead of quantitative intervals used in quantitative association rules, linguistic terms are used to represent the knowledge. As we can observe from the rule “If *middle* amount of *bread* is bought, then *high* amount of *milk* is bought,” *bread* and *milk* are items, and *middle* and *high* are linguistic terms. The rule means that if the

On Genetic-Fuzzy Data-Mining Techniques,

Fig. 2 A fuzzy association rule



quantity of the purchased item *bread* is *middle*, then there is a high possibility that the associated purchased item is *milk* with *high* quantity.

Many approaches have been proposed for mining fuzzy association rules (Chan and Au 1997; Hong et al. 1999; Kuok et al. 1998; Lee et al. 2004; Yue et al. 2000). Most of the approaches set a single minimum support threshold for all the items or itemsets and identify the association relationships among transactions. In real applications, different items may have different criteria to judge their importance. Multiple minimum support thresholds are thus proposed for this purpose. We can thus divide the fuzzy data-mining approaches into two types, namely Single-minimum-Support Fuzzy-Mining (*SSF**M*) (Chan and Au 1997; Hong et al. 1999; Kuok et al. 1998; Yue et al. 2000) and Multiple- minimum-Support Fuzzy-Mining (*MSF**M*) problems (Lee et al. 2004).

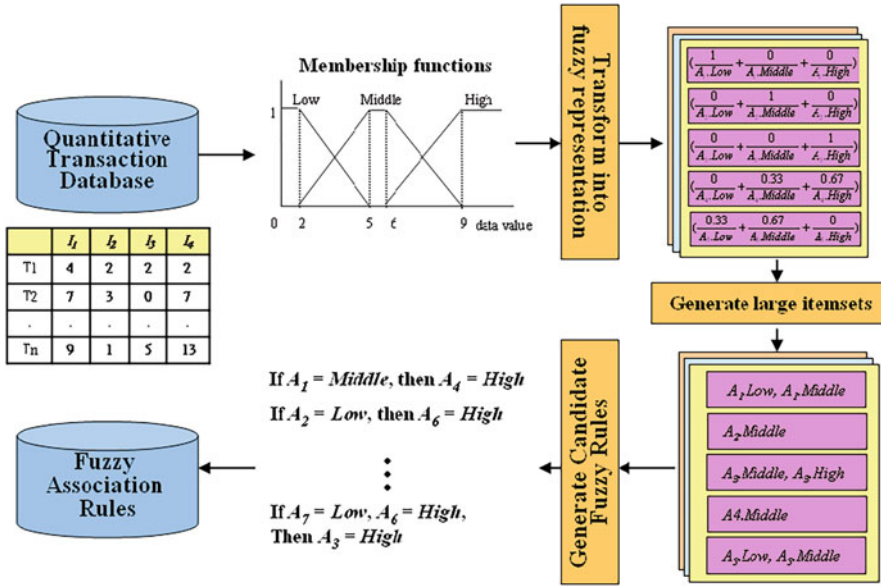
In the *SSF**M* problem, Chan and Au proposed an F-APACS algorithm to mine fuzzy association rules (Chan and Au 1997). They first transformed quantitative attribute values into linguistic terms and then used the adjusted difference analysis to find interesting associations among attributes. Kuok et al. proposed a mining approach for fuzzy association rules. Instead of minimum supports and minimum confidences used in most mining approaches, significance factors and certainty factors were used to derive large itemsets and fuzzy association rules (Kuok et al. 1998). At nearly the same time, Hong et al. proposed a fuzzy mining algorithm to mine fuzzy rules from quantitative transaction data (Hong et al. 1999). Basically, these fuzzy mining algorithms first used membership functions to transform each quantitative value into a fuzzy set in linguistic terms and

then used a fuzzy mining process to find fuzzy association rules. Yue et al. then extended the above concept to find fuzzy association rules with weighted items from transaction data (Yue et al. 2000). They adopted Kohonen self-organized mapping to derive fuzzy sets for numerical attributes. In general, the basic concept of fuzzy data mining for the *SSF**M* problem is shown in Fig. 3.

In Fig. 3, the process for fuzzy data mining first transforms quantity transactions into fuzzy representation according to the predefined membership functions. The transformed data are then calculated to generate large itemsets. Finally, the generated large itemsets are used to derive fuzzy association rules.

As to the *MSF**M* problem, Lee et al. proposed a mining algorithm which used multiple minimum supports to mine fuzzy association rules (Lee et al. 2004). They assumed that items had different minimum supports and the maximum constraint was used. That is, the minimum support for an itemset was set as the maximum of the minimum supports of the items contained in the itemset. Under the constraint, the characteristic of level-by-level processing was kept, such that the original Apriori algorithm could easily be extended to finding large itemsets. In addition to the maximum constraint, other constraints such as the minimum constraint can also be used with different rationale.

In addition to the above fuzzy mining approaches, fuzzy data mining with taxonomy or fuzzy taxonomy is developed. Fuzzy web mining is another application of it. Besides, fuzzy data mining is strongly related to fuzzy control, fuzzy clustering, and fuzzy learning.



On Genetic-Fuzzy Data-Mining Techniques, Fig. 3 The concept of fuzzy data mining for the SSFM problem

In most fuzzy mining approaches, the membership functions are usually predefined in advance. Membership functions are, however, very crucial to the final mined results. Below, we will describe how the genetic algorithm can be combined with fuzzy data mining to make the entire process more complete. The concept of the genetic algorithm will first be briefly introduced in the next section.

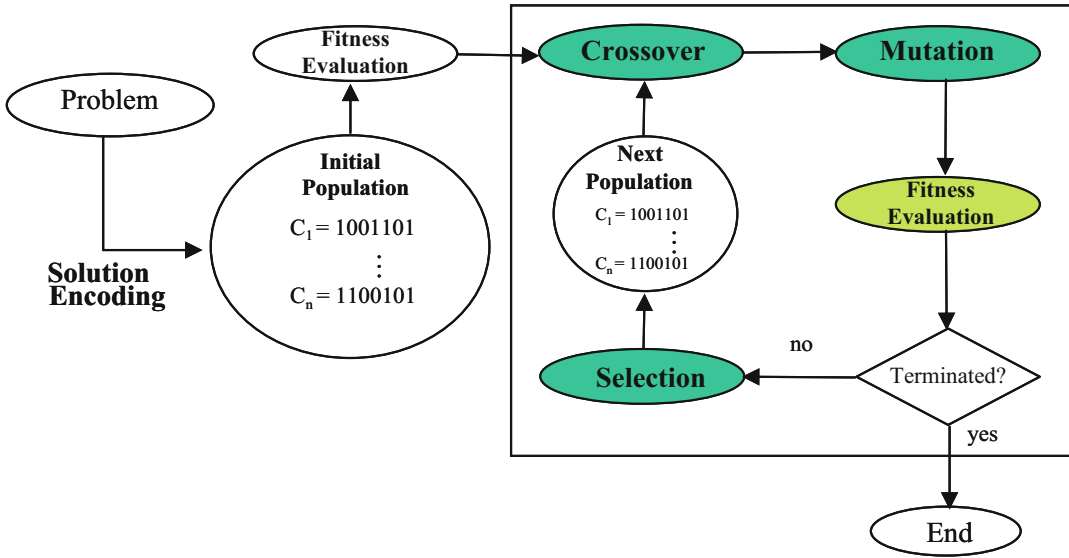
Genetic Algorithms

Genetic Algorithms (GAs) (Goldberg 1989; Holland 1975) have become increasingly important for researchers in solving difficult problems since they could provide feasible solutions in a limited amount of time (Homaifar et al. 1993). They were first proposed by Holland in 1975 (Holland 1975) and have been successfully applied to the fields of optimization (Goldberg 1989; Michalewicz 1994; Mitchell 1996), machine learning (Goldberg 1989; Michalewicz 1994), neural network (Mitchell 1996), fuzzy logic controllers (Sanchez et al. 1997), and so on. GAs are developed mainly based on the ideas and the techniques from genetic and evolutionary theory (Grefenstette 1986). According to

the principle of survival of the fittest, they generate the next population by several operations, with each individual in the population representing a possible solution. There are three principal operations in a genetic algorithm.

1. The *crossover* operation: It generates offspring from two chosen individuals in the population by exchanging some bits in the two individuals. The offspring thus inherit some characteristics from each parent.
2. The *mutation* operation: It generates offspring by randomly changing one or several bits in an individual. The offspring may thus possess different characteristics from their parents. Mutation prevents local searches of the search space and increases the probability of finding global optima.
3. The *selection* operation: It chooses some offspring for survival according to predefined rules. This keeps the population size within a fixed constant and puts good offspring into the next generation with a high probability.

On applying genetic algorithms to solving a problem, the first step is to define a representation that describes the problem states. The most



On Genetic-Fuzzy Data-Mining Techniques, Fig. 4 The entire GA process

common way used is the bit string representation. An initial population of *individuals*, called *chromosomes*, is then defined, and the three genetic operations (crossover, mutation, and selection) are performed to generate the next generation. Each chromosome in the population is evaluated by a *fitness function* to determine its goodness. This procedure is repeated until a user-specified termination criterion is satisfied. The entire GA process is shown in Fig. 4.

Genetic-Fuzzy Data-Mining Techniques

In the previous section for fuzzy data mining, several approaches were introduced, in which the membership functions were assumed to be known in advance. The given membership functions may, however, have a critical influence on the final mining results. Although many approaches for learning membership functions were proposed (Cordón et al. 2001; Roubos and Setnes 2001; Setnes and Roubos 2000; Wang et al. 1998, 2000), most of them were usually used for classification or control problems, and more details could be found in Herrera (2008). There were several strategies proposed for

learning membership functions in classification or control problems by genetic algorithms. Below are some of them:

1. Learning membership functions first, then rules
2. Learning rules first, then membership functions
3. Simultaneously learning rules and membership functions
4. Iteratively learning rules and membership functions

For fuzzy mining problems, many researches have also been done by combining the genetic algorithm and the fuzzy concepts to discover both suitable membership functions and useful fuzzy association rules from quantitative values. However, most of them adopt the first strategy. That is, the membership functions are first learned and then the fuzzy association rules are derived based on the obtained membership functions. It is done in this way because the number of association rules is often large in mining problems and cannot easily be coded in a chromosome.

In this entry, we introduce several genetic-fuzzy data-mining algorithms that can mine both

appropriate membership functions and fuzzy association rules (Chen et al. 2007a, b, 2008, 2009; Hong et al. 2004, 2006, 2008; Kaya and Alhaji 2005, 2006; Kaya 2006). The genetic-fuzzy mining problems can be divided into four kinds according to the types of fuzzy mining problems and the ways of processing items. The types of fuzzy mining problems include single-minimum-support fuzzy-mining (*SSFM*) and multiple-minimum-support fuzzy-mining (*MSFM*) as mentioned above. The ways of processing items include processing all the items together (integrated approach) and processing them individually (divide-and-conquer approach). The integrated genetic-fuzzy approaches encode all membership functions of all items (or attributes) into a chromosome (also called an individual). The genetic algorithms are then used to derive a set of appropriate membership functions according to the designed fitness function. Finally, the best set of membership functions are then used to mine fuzzy association rules. On the other hand, the divide-and-conquer genetic-fuzzy approaches encode membership functions of each item into a chromosome. In other words, chromosomes in a population were maintained just for only one item. The membership functions can thus be found for one item after another or at the same time via parallel processing. In general, the chromosomes in the divide-and-conquer genetic-fuzzy approaches are much shorter than those in the integrated approaches since the former only focus on individual items. But there are more application limitations on the former than on the latter. This will be explained later.

The four kinds of problems are thus the Integrated Genetic-Fuzzy problem for items with a Single Minimum Support (*IGFSMS*) (Chen et al. 2007a, 2008; Hong et al. 2006; Kaya and Alhaji 2005, 2006; Kaya 2006), the Integrated Genetic-Fuzzy problem for items with Multiple Minimum Supports (*IGFMMS*) (Chen et al. 2009), the Divide-and-Conquer Genetic-Fuzzy problem for items with a Single Minimum Support (*DGFSMS*) (Chen et al. 2007b, Hong et al. 2004, 2008), and the Divide-and-Conquer Genetic-Fuzzy problem for items with Multiple Minimum Supports (*DGFMMMS*). The classification is shown in Table 1.

On Genetic-Fuzzy Data-Mining Techniques, Table 1 The four different genetic-fuzzy data-mining problems

	Integrated approach	Divide-and-conquer approach
Single minimum support	<i>IGFSMS</i> problem	<i>DGFSMS</i> problem
Multiple minimum supports	<i>IGFMMS</i> problem	<i>DGFMMMS</i> problem

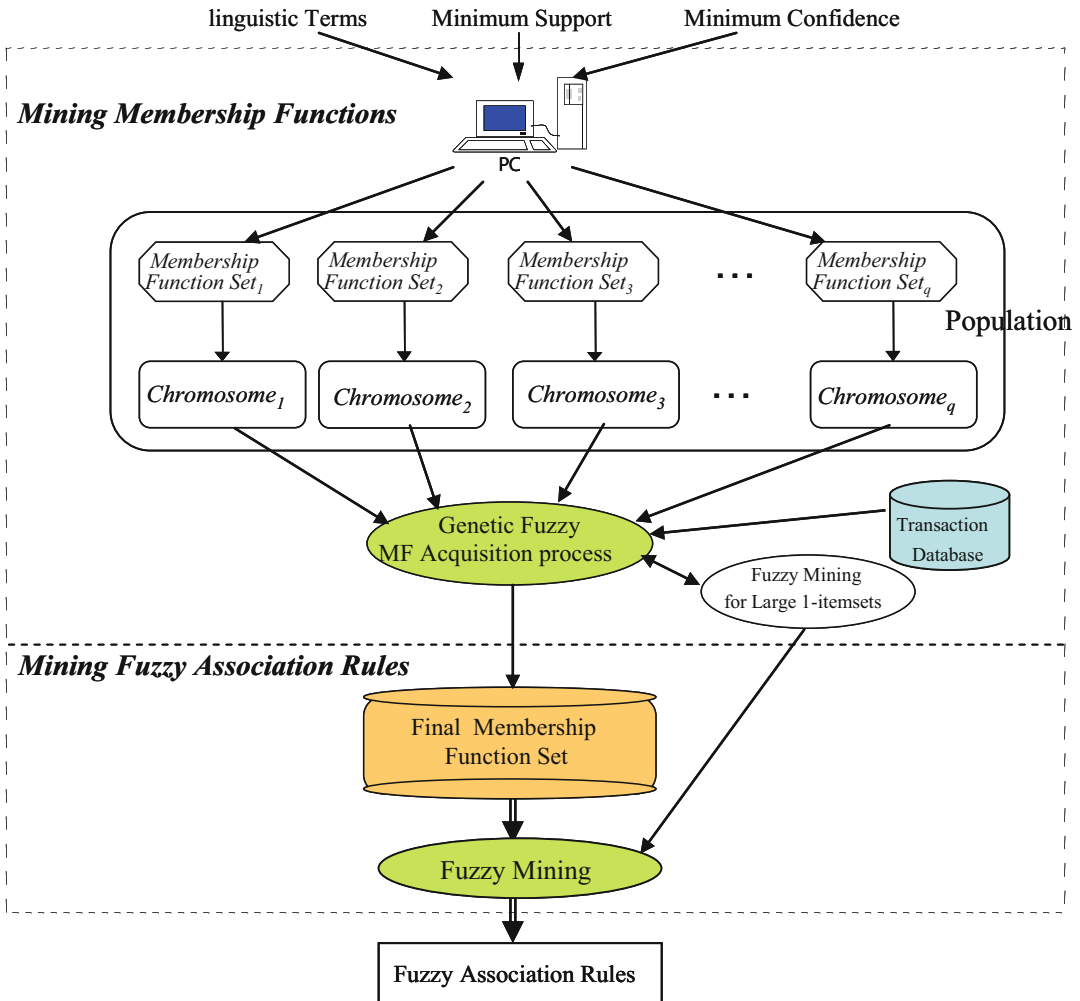
Each of the four kinds of genetic-fuzzy data-mining problems will be introduced in the following sections.

The Integrated Genetic-Fuzzy Problem for Items with a Single Minimum Support (*IGFSMS*)

Many approaches have been published for solving the *IGFSMS* problem (Chen et al. 2007a, 2008; Hong et al. 2006; Kaya and Alhaji 2005, 2006). For example, Hong et al. proposed a genetic-fuzzy data-mining algorithm for extracting both association rules and membership functions from quantitative transactions (Hong et al. 2006). They proposed a GA-based framework for searching membership functions suitable for given mining problems and then used the final best set of membership functions to mine fuzzy association rules. The proposed framework is shown in Fig. 5.

The proposed framework consists of two phases, namely mining membership functions and mining fuzzy association rules. In the first phase, the proposed framework maintains a population of sets of membership functions and uses the genetic algorithm to automatically derive the resulting one. It first transforms each set of membership functions into a fixed-length string. The chromosome is then evaluated by the number of large 1-itemsets and the suitability of membership functions. The fitness value of a chromosome C_q is then defined as:

$$f(C_q) = \frac{|L_1|}{\text{suitability}(C_q)},$$

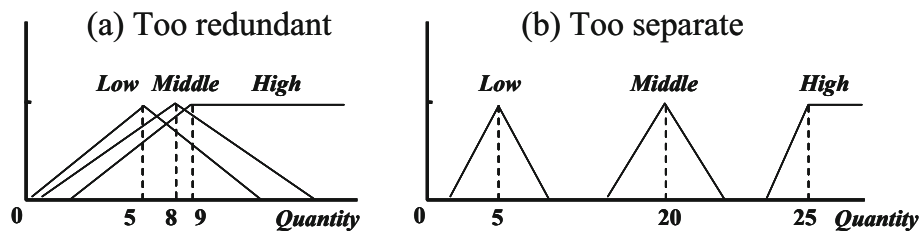


On Genetic-Fuzzy Data-Mining Techniques, Fig. 5 A genetic-fuzzy framework for the IGFSMS problem

where $|L_I|$ is the number of large 1-itemsets obtained by using the set of membership functions in C_q . Using the number of large 1-itemsets can achieve a trade-off between execution time and rule interestingness. Usually, a larger number of 1-itemsets will result in a larger number of all itemsets with a higher probability, which will thus usually imply more interesting association rules. The evaluation by 1-itemsets is, however, faster than that by all itemsets or interesting association rules. Of course, the number of all itemsets or interesting association rules can also be used in the fitness function. A discussion for different choices of fitness functions can be found in Chen et al. (2007a).

The suitability measure is used to reduce the occurrence of bad types of membership functions. The two bad types of membership functions are shown in Fig. 6, where the first one is too redundant, and the second one is too separate.

Two factors, called the overlap factor and the coverage factor, are used to avoid the bad shapes. The overlap factor is designed for avoiding the first bad case (too redundant), and the coverage factor is for the second one (too separate). Each factor has its formula for evaluating a value from a chromosome. After fitness evaluation, the approach then chooses appropriate chromosomes for mating, gradually creating good offspring membership function



On Genetic-Fuzzy Data-Mining Techniques, Fig. 6 The two bad types of membership functions

sets. The offspring membership function sets then undergo recursive evolution until a good set of membership functions has been obtained. In the second phase, the final best membership functions are gathered to mine fuzzy association rules. The fuzzy mining algorithm proposed in Hong et al. (2001) is adopted to achieve this purpose.

The calculation for large 1-itemsets, however, will still take a lot of time, especially when the database cannot totally be fed into the main memory. An enhanced approach, called the cluster-based fuzzy-genetic mining algorithm, was thus proposed (Chen et al. 2008) to speed up the evaluation process and keep nearly the same quality of solutions as that in Hong et al. (2006). That approach also maintains a population of sets of membership functions and uses the genetic algorithm to derive the best one. Before fitness evaluation, the clustering technique is first used to cluster chromosomes. It uses the *k*-means clustering approach to gather similar chromosomes into groups. The two factors, overlap factor and coverage factor, are used as two attributes for clustering. For example, assume there are ten chromosomes with their coverage and overlap factors shown in Table 2, where the column “suitability” represents the pair (coverage factor, overlap factor).

The *k*-means clustering approach is then executed to divide the ten chromosomes into *k* clusters. In this example, assume the parameter *k* is set at 3. The three clusters found are shown in Table 3. The representative chromosomes in the three clusters are *C*₅ (4.37, 0.33), *C*₄ (4.66, 0), and *C*₉ (4.09, 8.33).

All the chromosomes in a cluster use the number of large 1-itemsets derived from the

On Genetic-Fuzzy Data-Mining Techniques, Table 2 The coverage and the overlap factors of ten chromosomes

Chromosome	Suitability	Chromosome	Suitability
<i>C</i> ₁	(4, 0)	<i>C</i> ₆	(4.5, 0)
<i>C</i> ₂	(4.24, 0.5)	<i>C</i> ₇	(4.45, 0)
<i>C</i> ₃	(4.37, 0)	<i>C</i> ₈	(4.37, 0.53)
<i>C</i> ₄	(4.66, 0)	<i>C</i> ₉	(4.09, 8.33)
<i>C</i> ₅	(4.37, 0.33)	<i>C</i> ₁₀	(4.87, 0)

On Genetic-Fuzzy Data-Mining Techniques, Table 3 The three clusters found in the example

Cluster _{<i>i</i>}	Chromosomes	Representative chromosome
Cluster ₁	<i>C</i> ₁ , <i>C</i> ₂ , <i>C</i> ₅ , <i>C</i> ₈	<i>C</i> ₅
Cluster ₂	<i>C</i> ₃ , <i>C</i> ₄ , <i>C</i> ₆ , <i>C</i> ₇ , <i>C</i> ₁₀	<i>C</i> ₄
Cluster ₃	<i>C</i> ₉	<i>C</i> ₉

representative chromosome in the cluster and their own suitability of membership functions to calculate their fitness values. Since the number for scanning a database decreases, the evaluation cost can thus be reduced. In this example, the representative chromosomes are chromosomes *C*₄, *C*₅, and *C*₉, and it only needs to calculate the number of large 1-itemsets three times. The evaluation results are utilized to choose appropriate chromosomes for mating in the next generation. The offspring membership function sets then undergo recursive evolution until a good set of membership functions has been obtained. Finally, the

derived membership functions are used to mine fuzzy association rules.

Besides, Kaya and Alhaji also proposed several genetic-fuzzy data-mining approaches to derive membership functions and fuzzy association rules (Kaya and Alhaji 2005, 2006; Kaya 2006). In Kaya and Alhaji (2005), the proposed approach tries to derive membership functions, which can get a maximum profit within an interval of user-specified minimum support values. It then uses the derived membership functions to mine fuzzy association rules. The concept of their approaches is shown in Fig. 7.

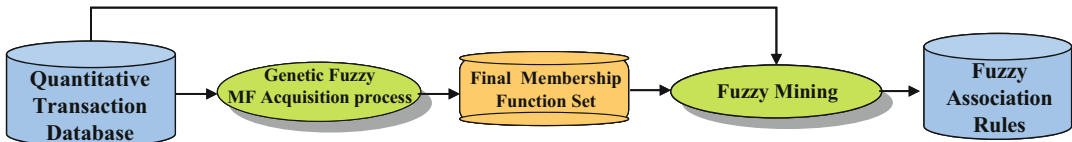
As shown in Fig. 7a, the approach first derives membership functions from the given quantitative transaction database by genetic algorithms. The final membership functions are then used to mine fuzzy association rules. Figure 7b shows the concept of maximizing the large itemsets of the given minimum support interval. It is used as the fitness function. Kaya and Alhaji also extended the approach to mine fuzzy weighted association rules (Kaya and Alhaji 2006). In addition, Matthews et al. took the temporal concept into consideration and proposed a temporal fuzzy

association rule mining with 2-tuples linguistic representation (Matthews et al. 2012).

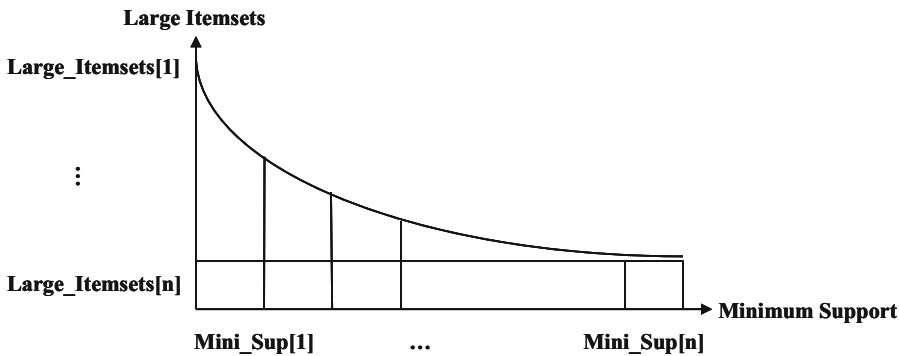
The above GFM algorithms focused on one single-concept level. In general, items may have taxonomy. Chen et al. thus proposed a multiple-level GFM algorithm for mining membership functions and fuzzy association rules on multiple-concept levels (Chen et al. 2012a). According to the given taxonomy, it first encodes the membership functions of each item class (category) into a chromosome. The fitness value of each individual is then evaluated by the summation of the large 1-itemsets of each item in different concept levels and the suitability of membership functions in each chromosome. After the GA process terminates, a better set of multiple-level fuzzy association rules can then be expected with a more appropriate set of membership functions.

The Integrated Genetic-Fuzzy Problem for Items with Multiple Minimum Supports (IGFMMS)

In the above subsection, it can be seen that lots of researches focus on integrated genetic-fuzzy



(a) The flowchart



(b) The concept of the designed fitness function

approaches for items with a single minimum support. However, different items may have different criteria to judge their importance. For example, assume among a set of items there are some which are expensive. They are thus seldom bought because of their high costs. Besides, the support values of these items are low. A manager may, however, still be interested in these products due to their high profits. In such cases, the above approaches may not be suitable for this problem. Chen et al. thus proposed another genetic-fuzzy data-mining approach (Chen et al. 2009), which was an extension of the approach proposed in Hong et al. (2006), to solve it. The approach combines the clustering, fuzzy, and genetic concepts to derive minimum support values and membership functions for items. The final minimum support values and membership functions are then used to mine fuzzy association rules. The genetic-fuzzy mining framework for the *IGFMMS* problem is shown in Fig. 8.

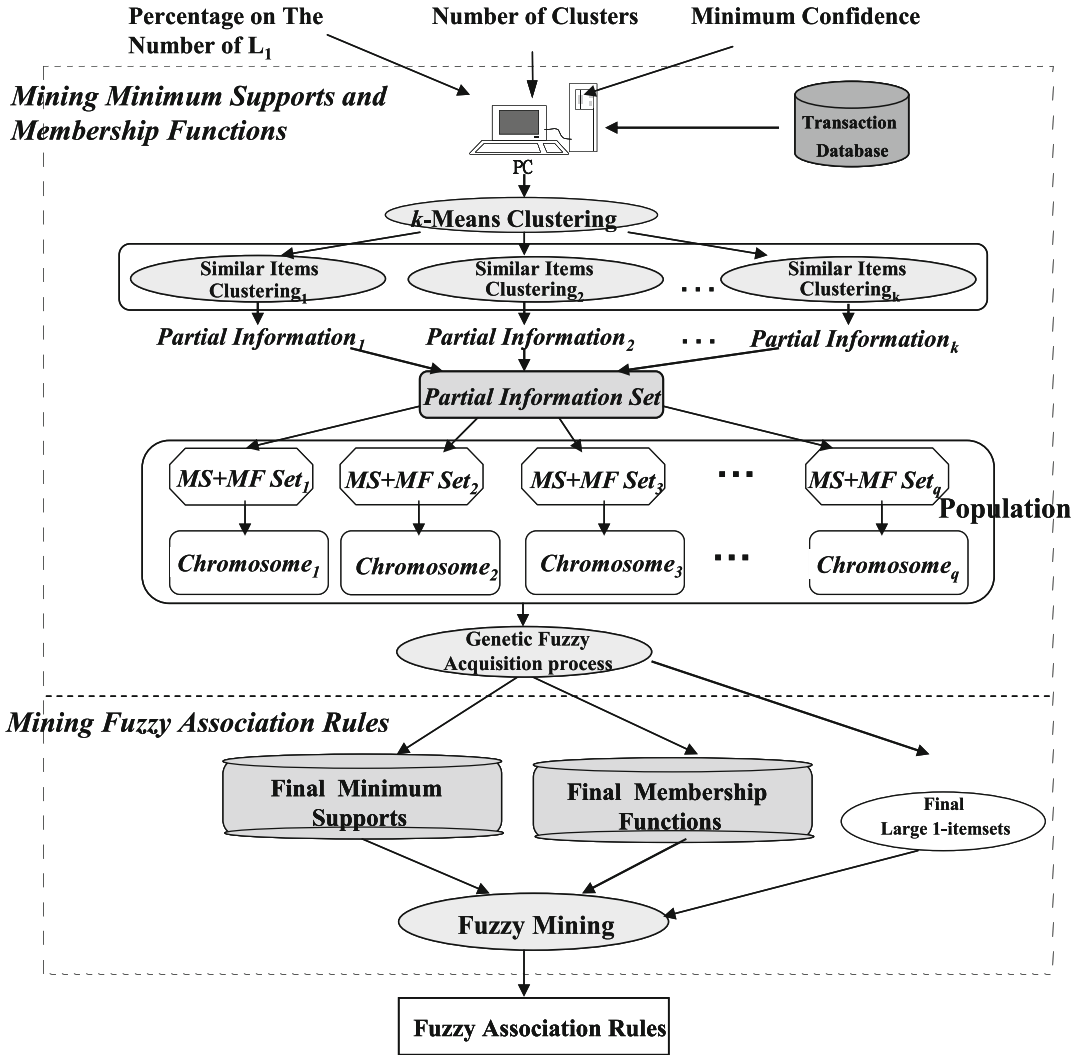
As shown in Fig. 8, the framework can be divided into two phases. The first phase searches for suitable minimum support values and membership functions of items, and the second phase uses the final best set of minimum support values and membership functions to mine fuzzy association rules. The proposed framework maintains a population of sets of minimum support values and membership functions and uses the genetic algorithm to automatically derive the resulting one.

A genetic algorithm requires a population of feasible solutions to be initialized and updated during the evolution process. As mentioned above, each individual within the population is a set of minimum support values and isosceles-triangular membership functions. Each membership function corresponds to a linguistic term of a certain item. In this approach, the initial set of chromosomes is generated based on the initialization information derived by the k-means clustering approach on the transactions. The frequencies and quantitative values of items in the transactions are the two main factors to gather similar items into groups. The initialization information includes an appropriate number of linguistic terms, the range of possible minimum support values, and

membership functions of each item. All the items in the same cluster are considered to have similar characteristics and are assigned similar initialization values when a population is initialized. The approach then generates and encodes each set of minimum support values and membership functions into a fixed-length string according to the initialization information.

In this approach, the minimum support values of items may be different. It is hard to assign the values. As an alternative, the values can be determined according to the required number of rules. It is, however, very time-consuming to obtain the rules for each chromosome. As mentioned above, a larger number of 1-itemsets will usually result in a larger number of all itemsets with a higher probability, which will thus usually imply more interesting association rules. The evaluation by 1-itemsets is faster than that by all itemsets or interesting association rules. Using the number of large 1-itemsets can thus achieve a trade-off between execution time and rule interestingness (Hong et al. 2006).

A criterion should thus be specified to reflect the user preference on the derived knowledge. In the approach, the required number of large 1-itemsets (*RNL*) is used for this purpose. It is the number of linguistic large 1-itemsets that a user wants to get from an item. It can be defined as the number of linguistic terms of an item multiplied by the predefined percentage which reflects users' preference on the number of large 1-itemsets. It is used to reflect the closeness degree between the number of derived large 1-itemsets and the required number of large 1-itemsets. For example, assume there are three linguistic terms for an item and the predefined percentage p is set at 80%. The *RNL* value is then set as $\lceil 3 \times 0.8 \rceil$, which is 2. The fitness function is then composed of the suitability of membership functions and the closeness to the *RNL* value. The minimum support values and membership functions can thus be derived by GA and are then used to mine fuzzy association rules by a fuzzy mining approach for multiple minimum supports such as the one in Lee et al. (2004).

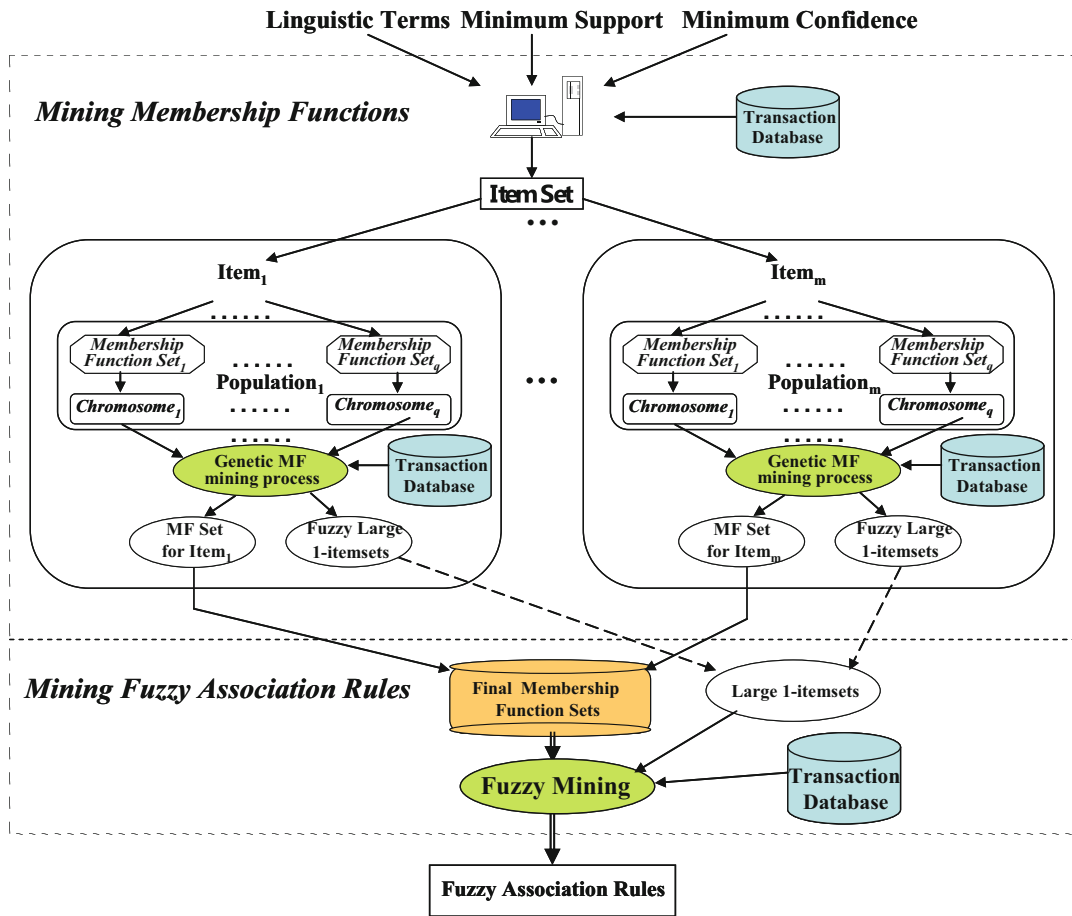


On Genetic-Fuzzy Data-Mining Techniques, Fig. 8 A genetic-fuzzy framework for the IGFMS problem

The Divide-and-Conquer Genetic-Fuzzy Problem for Items with a Single Minimum Support (DGFSMS)

The advantages of the integrated genetic-fuzzy approaches lie in that they are simple, easy to use, and with few constraints in the fitness functions. In addition to the number of large 1-itemsets, the other criteria can also be used. However, if the number of items is large, the integrated genetic-fuzzy approaches may need lots of time to find a near-optimal solution because the length of a chromosome is very long.

Recently, the divide-and-conquer strategy has been used in the evolutionary computation community to very good effect. Many algorithms based on it have also been proposed in different applications (Au et al. 2003; Darwen and Yao 1997; Khare et al. 2005; Yao 2003). When the number of large 1-itemsets is used in fitness evaluation, the divide-and-conquer strategy becomes a good choice to deal with it since each item can be individually processed in this situation. Hong et al. thus used a GA-based framework with the divide-and-conquer strategy to search for



On Genetic-Fuzzy Data-Mining Techniques, Fig. 9 A genetic-fuzzy framework for the DGFSMS problem

membership functions suitable for the mining problem (Hong et al. 2008). The framework is shown in Fig. 9.

The proposed framework in Fig. 9 is divided into two phases: mining membership functions and mining fuzzy association rules. Assume the number of items is m . In the phase of mining membership functions, it maintains m populations of membership functions, with each population for an item. Each chromosome in a population represents a possible set of memberships for that item. The chromosomes in the same population are of the same length. The fitness of each set of membership functions is evaluated by the fuzzy-supports of the linguistic terms in large 1-itemsets and by the suitability of the derived membership functions.

The offspring sets of membership functions undergo recursive evolution until a good set of membership functions has been obtained. Next, in the phase of mining fuzzy association rules, the sets of membership function for all the items are gathered together and used to mine the fuzzy association rules from the given quantitative database. Besides, an enhanced approach (Chen et al. 2007b), which combines the clustering and the divide-and-conquer techniques, was also proposed to speed up the evaluation process. The clustering idea is similar to that in IGFSMS (Chen et al. 2008) except that the center value of each membership function is also used as an attribute to cluster chromosomes. The clustering process is thus executed according to the coverage factors, the overlap factors, and the center

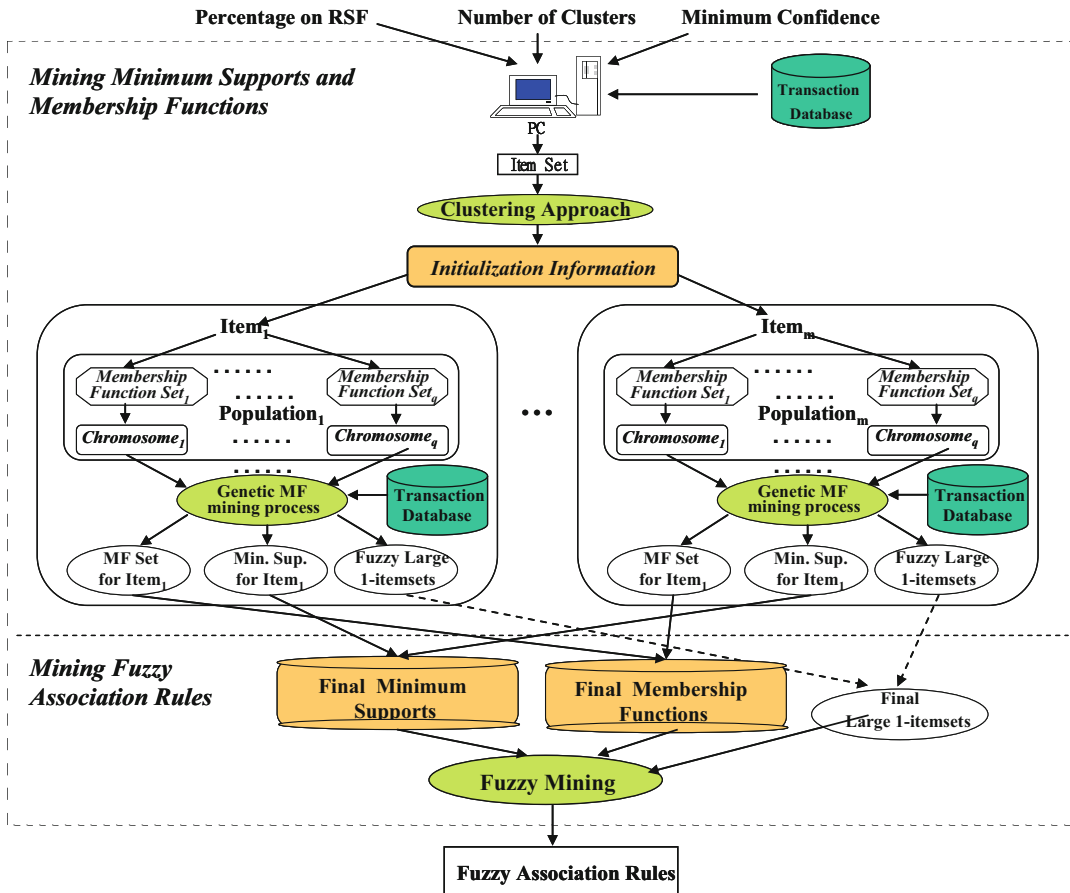
values of chromosomes. For example, assume each item has three membership functions. Totally five attributes including one coverage factor, one overlap factor, and three center values are used to form appropriate clusters. Note that the number of linguistic terms for each item is predefined in the mentioned approaches. It may also be automatically and dynamically adjusted (Hong et al. 2004).

The Divide-and-Conquer Genetic-Fuzzy Problem for Items with Multiple Minimum Supports (DGFMMS)

The problem may be thought of as the combination of the *IGFMMS* and the *DGFSMS* problems. The framework for the *DGFMMS* problem can thus be easily designed from the previous

frameworks for *IGFMMS* and *DGFSMS*. It is shown in Fig. 10.

The proposed framework in Fig. 10 is divided into two phases: mining minimum supports and membership functions, and mining fuzzy association rules. In the first phase, the clustering approach is first used for deriving initialization information which is then used for getting better initial populations as used for *IGFMMS*. It then maintains m populations of minimum supports and membership functions, with each population for an item. Next, in the phase of mining fuzzy association rules, the minimum support values and membership functions for all the items are gathered together and are used to mine fuzzy interesting association rules from the given quantitative database.



On Genetic-Fuzzy Data-Mining Techniques, Fig. 10 A genetic-fuzzy framework for the *DGFMMS* problem

Multiobjective Genetic-Fuzzy Data Mining

Furthermore, since single objective function may not be good enough for dealing with a real world application, the multiobjective genetic algorithms have been developed for solving such problems (Ishibuchi and Yamamoto 2004; Jin 2006). In other words, more than one criterion are used in the evaluation. A set of solutions, namely non-dominated points (also called Pareto-Optimal Surface), is derived and given to users, instead of only the best one solution obtained by genetic algorithms. Kaya and Alhaji thus proposed an approach based on multiobjective genetic algorithms to learn membership functions, which were then used to generate interesting fuzzy association rules (Kaya 2006). Three objective functions, namely strongness, interestingness, and comprehensibility, were used in their approach to find the Pareto-Optimal Surface. Hong et al. adopted a sophisticated multiobjective approach, the SPEA2, to find appropriate sets of membership functions for fuzzy data mining. Two objective functions were used to find the Pareto front. The first one was the suitability of membership functions, and the second one was the total number of large 1-itemsets derived (Chen et al. 2012b). Matthews et al. then utilized three objectives including temporal support, fuzzy itemset support, and membership function widths for deriving temporal fuzzy association rules (Matthews et al. 2011).

As mentioned in the previous section, items may have taxonomy. Based on the algorithm presented in (Chen et al. 2012a), Hong et al. have proposed a Multi-Objective Multiple-Level GFM algorithm (MOMLGFM) for mining sets of membership functions (Pareto solutions) and fuzzy association rules on multiple-concept levels. The proposed framework is shown in Fig. 11.

The proposed framework consists of two phases that are multiobjective genetic-fuzzy MF acquisition process and multilevel fuzzy association rule mining process. In the first phase, it first generates and transforms membership functions of each item category into a fixed-length string

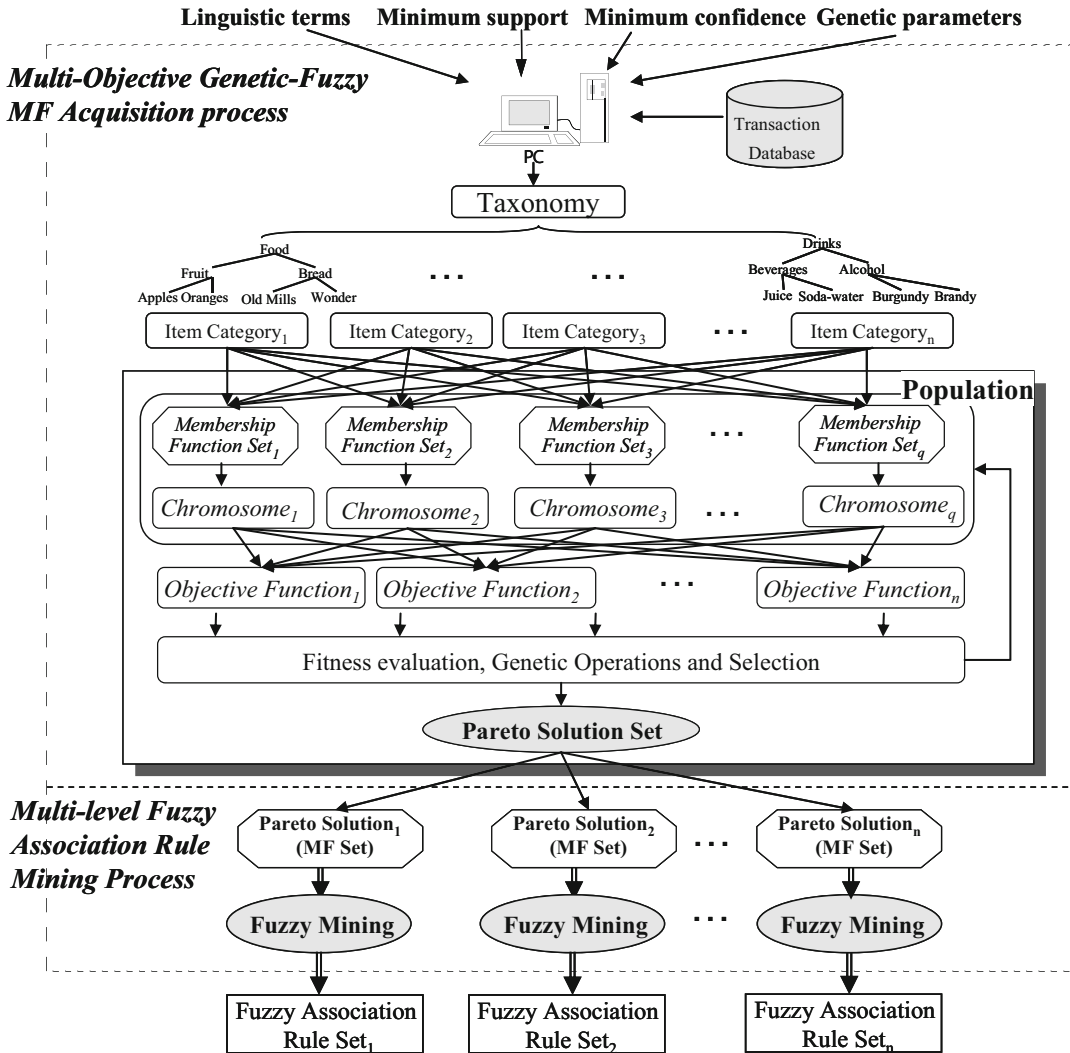
which is known as a chromosome or an individual according to the given taxonomy. The values of objective functions of all chromosomes are then calculated for evaluating fitness values and executing genetic operations. The Pareto solution set is then used to keep the nondominated solutions during the acquisition process. In the second phase, each Pareto solution (a set of membership functions) is used to mine multiple-level fuzzy association rules. Note that various rule mining approaches could be used in this phase. A detailed review of the application of multiobjective evolutionary systems could be found in Fazzolari et al. (2013).

Future Directions

In this entry, we have introduced some genetic-fuzzy data-mining techniques and their classification. The concept of fuzzy sets is used to handle quantitative transactions, and the process of genetic calculation is executed to find appropriate membership functions. The genetic-fuzzy mining problems are divided into four kinds according to the types of fuzzy mining problems and the ways of processing items. The types of fuzzy mining problems include single-minimum-support fuzzy-mining (SSFM) and multiple-minimum-support fuzzy-mining (MSFM). The ways of processing items include processing all the items together (integrated approach) and processing them individually (divide-and-conquer approach). Each of the four kinds of problems has been described with some approaches given.

Data mining is very important especially when the data amounts in the information era are extremely large. The topic will continuously grow but with a variety of forms. Some possible research directions in the future about genetic-fuzzy data mining are listed as follows.

- *Analyzing effects of different shapes of membership functions and different genetic operators:* Different shapes of membership functions may have different results on genetic-fuzzy data mining. They may be



On Genetic-Fuzzy Data-Mining Techniques, Fig. 11 The MOGA-based multilevel fuzzy-data-mining framework

evaluated in the future. Different genetic operations may also be tried to get better results than the ones used in the above approaches.

- Enhancing performance of the fuzzy-rule mining phase:** The final goal of genetic-fuzzy mining techniques introduced in this entry is to mine appropriate fuzzy association rules. It is, however, very time-consuming in the phase of mining fuzzy association rules. How to improve the process of mining interesting fuzzy rules is thus worth studying. Some possible approaches include modifying existing

approaches, combining the existing ones with other techniques, or defining new evaluation criteria.

- Developing visual tools for these genetic-fuzzy mining approaches:** Another interesting future work is to develop visual tools for demonstrating the genetic-fuzzy mining results. It can help a decision maker easily understand or get useful information quickly. The visual tools may include, for example, how to show the derived membership functions and to illustrate the interesting fuzzy association rules.

- **Designing genetic-fuzzy mining approaches for dealing with large-scale datasets:** In recent years, the data size increases sharply due to the quick growth of equipments and wireless networks. How to design genetic-fuzzy mining approaches for dealing with large-scale and high-dimensional datasets becomes a challenge. The MapReduce technique which is a powerful framework recently developed in cloud environments may be utilized for solving these issues (Jiang et al. 2011).

Bibliography

- Agrawal R, Srikant R (1994) Fast algorithm for mining association rules. In: The international conference on very large data bases, pp 487–499
- Agrawal R, Imielinski T, Swami A (1993a) Mining association rules between sets of items in large database. In: The 1993 ACM SIGMOD conference, Washington, DC
- Agrawal R, Imielinski T, Swami A (1993b) Database mining: a performance perspective. *IEEE Trans Knowl Data Eng* 5(6):914–925
- Agrawal R, Srikant R, Vu Q (1997) Mining association rules with item constraints. In: The third international conference on knowledge discovery in databases and data mining, Newport Beach, California, August 1997
- Au WH, Chan KCC, Yao X (2003) A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Trans Evol Comput* 7(6):532–545
- Aumann Y, Lindell Y (1999) A statistical theory for quantitative association rules. In: The ACM SIGKDD international conference on knowledge discovery and data mining, pp 261–270
- Casillas J, Cordon O, del Jesus MJ, Herrera F (2005) Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans Fuzzy Syst* 13(1):13–29
- Chan CC, Au WH (1997) Mining fuzzy association rules. In: The conference on information and knowledge management, Las Vegas, pp 209–215
- Chen J, Mikulcic A, Kraft DH (2000) An integrated approach to information retrieval with fuzzy clustering and fuzzy inferencing. In: Pons O, Vila MA, Kacprzyk J (eds) Knowledge management in fuzzy databases. Physica-Verlag, Heidelberg
- Chen CH, Hong TP, Tseng VS (2007a) A comparison of different fitness functions for extracting membership functions used in fuzzy data mining. In: IEEE symposium on foundations of computational intelligence, pp 550–555
- Chen CH, Hong TP, Tseng VS (2007b) A modified approach to speed up genetic-fuzzy data mining with divide-and-conquer strategy. In: The IEEE congress on evolutionary computation (CEC), pp 1–6
- Chen CH, Tseng VS, Hong TP (2008) Cluster-based evaluation in fuzzy-genetic data mining. *IEEE Trans Fuzzy Syst* 16(1):249–262
- Chen CH, Hong TP, Tseng VS, Lee CS (2009) A genetic-fuzzy mining approach for items with multiple minimum supports. *Soft Comput* 13(5):521–533
- Chen CH, Hong TP, Lee YC (2012a) Genetic-fuzzy mining with taxonomy. *Int J Uncertainty Fuzziness Knowledge Based Syst* 20(2):187–205
- Chen CH, Hong TP, Tseng VS (2012b) Finding Pareto-front membership functions in fuzzy data mining. *Int J Comput Intell Syst* 5(2):343–354
- Cordon O, Herrera F, Villar P (2001) Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Trans Fuzzy Syst* 9(4):667–674
- Darwen PJ, Yao X (1997) Speciation as automatic categorical modularization. *IEEE Trans Evol Comput* 1(2):101–108
- Fazzolari M, Alcalá R, Nojima Y, Ishibuchi H, Herrera F (2013) A review of the application of multi-objective evolutionary systems: current status and further directions. *IEEE Trans Fuzzy Syst* 21(1):45–65
- Frawley WJ, Piatetsky-Shapiro G, Matheus CJ (1991) Knowledge discovery in databases: an overview. In: The AAAI workshop on knowledge discovery in databases, pp 1–27
- Goldberg DE (1989) Genetic algorithms in search, optimization & machine learning. Addison Wesley, Reading
- Grefenstette JJ (1986) Optimization of control parameters for genetic algorithms. *IEEE Trans Syst Man Cybern* 16(1):122–128
- Heng PA, Wong TT, Rong Y, Chui YP, Xie YM, Leung KS, Leung PC (2006) Intelligent inferencing and haptic simulation for Chinese acupuncture learning and training. *IEEE Trans Inf Technol Biomed* 10(1):28–41
- Herrera F (2008) Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evol Intel* 1:27–46
- Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
- Homaifar A, Guan S, Liepins GE (1993) A new approach on the traveling salesman problem by genetic algorithms. In: The fifth international conference on genetic algorithms
- Hong TP, Lee YC (2001) Mining coverage-based fuzzy rules by evolutionary computation. In: The IEEE international conference on data mining, pp 218–224
- Hong TP, Kuo CS, Chi SC (1999) Mining association rules from quantitative data. *Intell Data Anal* 3(5):363–376
- Hong TP, Kuo CS, Chi SC (2001) Trade-off between time complexity and number of rules for fuzzy mining from quantitative data. *Int J Uncertainty Fuzziness Knowledge Based Syst* 9(5):587–604
- Hong TP, Chen CH, Wu YL, Tseng VS (2004) Finding active membership functions in fuzzy data mining. In: The workshop on foundations of data mining in the fourth IEEE international conference on data mining
- Hong TP, Chen CH, Wu YL, Lee YC (2006) A GA-based fuzzy mining approach to achieve a trade-off between

- number of rules and suitability of membership functions. *Soft Comput* 10(11):1091–1101
- Hong TP, Chen CH, Lee YC, Wu YL (2008) Genetic-fuzzy data mining with divide-and-conquer strategy. *IEEE Trans Evol Comput* 12(2):252–265
- Ishibuchi H, Yamamoto T (2004) Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets Syst* 141:59–88
- Ishibuchi H, Yamamoto T (2005) Rule weight specification in fuzzy rule-based classification systems. *IEEE Trans Fuzzy Syst* 13(4):428–435
- Jiang D, Tung AKH, Chen G (2011) MAP-JOIN-REDUCE: toward scalable and efficient data analysis on large clusters. *IEEE Trans Knowl Data Eng* 23(9):1299–1311
- Jin Y (2006) *Multi-objective machine learning*. Springer, New York
- Kaya M (2006) Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules. *Soft Comput* 10(7):578–586
- Kaya M, Alhajj R (2005) Genetic algorithm based framework for mining fuzzy association rules. *Fuzzy Sets Syst* 152(3):587–601
- Kaya M, Alhajj R (2006) Utilizing genetic algorithms to optimize membership functions for fuzzy weighted association rules mining. *Appl Intell* 24(1):7–15
- Khare VR, Yao X, Sendhoff B, Jin Y, Wersing H (2005) Co-evolutionary modular neural networks for automatic problem decomposition. In: *The 2005 IEEE congress on evolutionary computation*, vol 3, pp 2691–2698
- Kuok C, Fu A, Wong M (1998) Mining fuzzy association rules in databases. *SIGMOD Rec* 27(1):41–46
- Lee YC, Hong TP, Lin WY (2004) Mining fuzzy association rules with multiple minimum supports using maximum constraints. *Lect Notes Comput Sci* 3214:1283–1290
- Liang H, Wu Z, Wu Q (2002) A fuzzy based supply chain management decision support system. *World Congress Intell Control Autom* 4:2617–2621
- Mamdani EH (1974) Applications of fuzzy algorithms for control of simple dynamic plants. *IEEE Proc* 121:1585–1588
- Matthews SG, Gongora MA, Hopgood AA (2011) Evolving temporal fuzzy itemsets from quantitative data with a multi-objective evolutionary algorithm. In: *The IEEE international workshop on genetic and evolutionary fuzzy systems*, pp 9–16
- Matthews SG, Gongora MA, Hopgood AA, Ahmadi S (2012) Temporal fuzzy association rule mining with 2-tuple linguistic representation. In: *The IEEE international conference on fuzzy systems*, pp 1–8
- Michalewicz Z (1994) *Genetic algorithms + data structures = evolution programs*. Springer, New York
- Mitchell M (1996) *An introduction to genetic algorithms*. MIT press, Cambridge, MA
- Rasmani KA, Shen Q (2004) Modifying weighted fuzzy subthreshold-based rule models with fuzzy quantifiers. *IEEE Int Conf Fuzzy Syst* 3:1679–1684
- Roubos H, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Trans Fuzzy Syst* 9(4):516–524
- Sanchez E, et al (1997) *Genetic algorithms and fuzzy logic systems: soft computing perspectives*. *Advances in fuzzy systems – applications and theory*, vol 7. World-Scientific
- Setnes M, Roubos H (2000) GA-fuzzy modeling and classification: complexity and performance. *IEEE Trans Fuzzy Syst* 8(5):509–522
- Siler W, James J (2004) *Fuzzy expert systems and fuzzy reasoning*. Wiley
- Srikant R, Agrawal R (June 1996) Mining quantitative association rules in large relational tables. In: *The 1996 ACM SIGMOD international conference on management of data*, Monreal, Canada, pp 1–12
- Wang CH, Hong TP, Tseng SS (1998) Integrating fuzzy knowledge by genetic algorithms. *IEEE Trans Evol Comput* 2(4):138–149
- Wang CH, Hong TP, Tseng SS (2000) Integrating membership functions and fuzzy rule sets from multiple knowledge sources. *Fuzzy Sets Syst* 112:141–154
- Yao X (2003) Adaptive divide-and-conquer using populations and ensembles. In: *The 2003 international conference on machine learning and application*, pp 13–20
- Yue S, Tsang E, Yeung D, Shi D (2000) Mining fuzzy association rules with weighted items. In: *Proceedings of the international IEEE conference on systems, man and cybernetics*, pp 1906–1911
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zhang H, Liu D (2006) *Fuzzy modeling and fuzzy control*. Springer
- Zhang Z, Lu Y, Zhang B (1997) An effective partitioning-combining algorithm for discovering quantitative association rules. In: *The Pacific-Asia conference on knowledge discovery and data mining*, pp 261–270

Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach

Salvatore Greco¹, Benedetto Matarazzo¹ and Roman Słowiński^{2,3}

¹Department of Economics and Business, University of Catania, Catania, Italy

²Institute of Computing Science, Poznań University of Technology, Poznań, Poland

³Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction: Granular Computing and

Ordered Data

Philosophical Basis of DRSA Granular

Computing

Dominance-Based Rough Set Approach

Fuzzy Set Extensions of the Dominance-Based Rough Set Approach

Variable-Consistency Dominance-Based Rough Set Approach (VC-DRSA)

Dominance-Based Rough Approximation of a Fuzzy Set

Monotonic Rough Approximation of a Fuzzy Set Versus Classical Rough Set

Dominance-Based Rough Set Approach to Case-Based Reasoning

An Algebraic Structure for Dominance-Based Rough Set Approach

Conclusions

Future Directions

Bibliography

Glossary

Granular Computing Granular computing is a general computation theory for using granules

such as subsets, classes, objects, clusters, and elements of a universe to build an efficient computational model for complex applications with huge amounts of data, information, and knowledge. Granulation of an object a leads to a collection of granules, with a granule being a clump of points (objects) drawn together by indiscernibility, similarity, proximity, or functionality. In human reasoning and concept formulation, the granules and the values of their attributes are fuzzy rather than crisp. In this perspective, fuzzy information granulation may be viewed as a mode of generalization, which can be applied to any concept, method, or theory.

Fuzzy Sets Differently from ordinary sets in which an object belongs or does not belong to a given set, in a fuzzy set an object belongs to a set in some degree. Formally, in universe U a fuzzy set X is characterized by its membership function $\mu_X : U \rightarrow [0, 1]$, such that for any $y \in U$, y certainly does not belong to set X if $\mu_X(y) = 0$, y certainly belongs to X if $\mu_X(y) = 1$, and y belongs to X with a given degree of certainty represented by the value of $\mu_X(y)$ in all other cases.

Rough Set A rough set in universe U is an approximation of a set based on available information about objects of U . The rough approximation is composed of two ordinary sets called *lower and upper approximation*. Lower approximation is a maximal subset of objects which, according to the available information, certainly belong to the approximated set, and upper approximation is a minimal subset of objects which, according to the available information, possibly belong to the approximated set. The difference between upper and lower approximation is called *boundary*.

Decision Rule Decision rule is a logical statement of the type “if..., then...,” where the premise (condition part) specifies values assumed by one or more condition attributes and the conclusion (decision part) specifies an overall judgment.

Ordinal Properties and Monotonicity Ordinal properties in description of objects are related to graduality of the presence or absence of a property. In this context, it is meaningful to say that a property is more present in one object than in another object. It is important that the ordinal descriptions are handled properly, which means, without introducing any operation, such as sum, averages, or fuzzy operators, like t-norm or t-conorm of Łukasiewicz, taking into account cardinal properties of data not present in the considered descriptions, which would, therefore, give not meaningful results. Monotonicity is strongly related to ordinal properties. It regards relationships between degrees of presence and absence of properties in the objects, like “the more present is property f_i , the more present is property f_j ,” or “the more present is property f_i , the more absent is property f_j .” The graded presence or absence of a property can be meaningfully represented using fuzzy sets. More precisely, the degree of presence of property f_i in object $y \in U$ is the value given to y by the membership function of the set of objects having property f_i .

Dominance-Based Rough Set Approach (DRSA) DRSA permits approximation of a set in universe U based on available ordinal information about objects of U . Also the decision rules induced within DRSA are based on ordinal properties of the elementary conditions in the premise and in the conclusion, such as “*if property f_{i1} is present in degree at least α_{i1} and ... property f_{ip} is present in degree at least α_{ip} , then property f_{iq} is present in degree at least α_{iq}* ”

Case-Based Reasoning Case-based reasoning is a paradigm in machine learning whose idea is that a new problem can be solved by noticing its similarity to a set of problems previously solved. Case-based reasoning regards the inference of some proper conclusions related to a new situation by the analysis of similar cases from a memory of previous cases. Very often similarity between two objects is expressed on a graded scale, and this justifies application of fuzzy sets in this context. Fuzzy case-based reasoning is a popular approach in this domain.

Definition of the Subject and Its Importance

This article describes dominance-based rough set approach (DRSA) to granular computing and data mining. DRSA was first introduced as a generalization of the rough set approach for dealing with multicriteria decision analysis, where preference order is important. The ordering is also important, however, in many other problems of data analysis. Even when the ordering seems absent, the presence or the absence of a property can be represented in ordinal terms, because if two properties are related, the presence, rather than the absence, of one property should make more (or less) probable the presence of the other property. This is even more apparent when the presence or the absence of a property is graded or fuzzy, because in this case, the more credible the presence of a property, the more (or less) probable the presence of the other property. Since the presence of properties, possibly fuzzy, is the basis of any granulation, DRSA can be seen as a general basis for granular computing.

After presenting the main ideas of DRSA for granular computing and its philosophical basis, the article introduces the basic concepts of DRSA, followed by its extensions in a fuzzy context and in probabilistic terms. This prepares the ground for treating the rough approximation of a fuzzy set, which is the core of the subject. It is also explained why classical rough set approach is a specific case of DRSA. The article continues with presentation of DRSA for case-based reasoning, where the main ideas of DRSA for granular computing are fruitfully applied. Finally, some basic formal properties of the whole approach are presented in terms of an algebra modeling the logic of DRSA.

Introduction: Granular Computing and Ordered Data

Granular computing originated in the research of Lin (1988, 1989, 1992, 1997, 1998a, b) and Zadeh (1979, 1996, 1997, 1999) and gained considerable interest in the last decade. The basic components

of granular computing are granules, such as subsets, classes, objects, clusters, and elements of a universe. Granulation of an object a leads to a collection of granules, with a granule being a clump of points (objects) drawn together by indiscernibility, similarity, proximity, or functionality. In human reasoning and concept formulation, the granules and the values of their attributes are fuzzy rather than crisp. In this perspective, fuzzy information granulation may be viewed as a mode of generalization, which can be applied to any concept, method or theory. Moreover, the theory of fuzzy granulation provides a basis for computing with words, due to the observation that in a natural language, words play the role of labels of fuzzy granules.

Since fuzzy granulation plays a central role in fuzzy logic and in its applications, and rough set theory can be considered as a crisp granulation of set theory, it is interesting to study the relationship between fuzzy sets and rough sets from this point of view. Moreover, noticing that fuzzy granulation that leads to fuzzy logic underlies all applications of granulation, hybridization of fuzzy sets and rough sets can lead to a more general theory of granulation with a potential of application to any domain of human investigation. This explains the interest in putting together rough sets and fuzzy sets.

Recently, it has been shown that a proper way of handling graduality in rough set theory is to use the dominance-based rough set approach (DRSA) (Greco et al. 2006b). This implies that DRSA is also a proper way of handling granulation within rough set theory. Let us explain this point in detail.

Rough set approach has been proposed to approximate some relationships existing between concepts. For example, in medical diagnosis the concept of “disease Y” can be represented in terms of such concepts as “low blood pressure” and “high temperature” or “muscle pain” and “headache.” The classical rough approximation is based on a very coarse representation, that is, for each aspect characterizing a concept (“low blood pressure,” “high temperature,” “muscle pain,” etc.), only its presence or its absence is considered relevant. In this case, the rough approximation involves a very primitive idea of monotonicity related to a scale with only two values: “presence” and “absence.”

Monotonicity gains importance when a finer representation of the concepts is considered. A representation is finer when, for each aspect characterizing a concept, not only its presence or its absence is taken into account, but also the *degree* of its presence or absence is considered relevant. Graduality is typical for fuzzy set philosophy (Zadeh 1965) and, therefore, a joint consideration of rough sets and fuzzy sets is worthwhile. In fact, rough sets and fuzzy sets capture the two basic complementary aspects of monotonicity: rough sets deal with relationships between different concepts, and fuzzy sets deal with expression of different dimensions in which the concepts are considered. For this reason, many approaches have been proposed to combine fuzzy sets with rough sets (see, e.g., Dubois and Prade (1990, 1992b), Pal and Skowron (1999), Radzikowska and Kerre (2002), Peters et al. (2005), and Słowiński et al. (2014)). Our combination of rough sets and fuzzy sets presents some important advantages with respect to the other approaches, which are discussed below.

The main preoccupation in almost all the studies combining rough sets with fuzzy sets was related to a fuzzy extension of Pawlak’s definition of lower and upper approximations using fuzzy connectives (Fodor and Roubens 1994; Klement et al. 2000). In fact, there is no rule for the choice of the “right” connective, so this choice is always arbitrary to some extent. Another drawback of fuzzy extensions of rough sets involving fuzzy connectives is that they are based on cardinal properties of membership degrees. In consequence, the result of these extensions is sensitive to order preserving transformation of membership degrees. For example, consider the t-conorm of Łukasiewicz as fuzzy connective; it may be used in the definition of both fuzzy lower approximation (to build fuzzy implication) and fuzzy upper approximation (as a fuzzy counterpart of a union). The t-conorm of Łukasiewicz is defined as

$$T^*(\alpha, \beta) = \min(\alpha + \beta, 1), \quad \alpha, \beta \in [0, 1].$$

$T^*(\alpha, \beta)$ can be interpreted as follows. Given two fuzzy propositions p and q , putting $v(p) = \alpha$ and $v(q) = \beta$, $T^*(\alpha, \beta)$ can be interpreted as

$v(p \vee q)$, the truth value of the proposition $p \vee q$. Let us consider the following values of arguments:

$$\alpha = 0.5, \quad \beta = 0.3, \quad \gamma = 0.2, \quad \delta = 0.1,$$

and their order preserving transformation:

$$\alpha' = 0.4, \quad \beta' = 0.3, \quad \gamma' = 0.2, \quad \delta' = 0.05.$$

The values of the t-conorm are in the two cases as follows:

$$\begin{aligned} T^*(\alpha, \delta) &= 0.6, & T^*(\beta, \gamma) &= 0.5, & T^*(\alpha', \delta') \\ &= 0.45, & T^*(\beta', \gamma') &= 0.5. \end{aligned}$$

One can see that the order of the results has changed after the order preserving transformation of the arguments. This means that the Łukasiewicz t-conorm takes into account not only the ordinal properties of the truth values but also their cardinal properties. A natural question arises: is it reasonable to expect from truth values a cardinal content instead of ordinal only? Or, in other words, is it realistic to claim that a human is able to say in a meaningful way not only that

- (a) “proposition p is more credible than proposition q ” but even something like
- (b) “proposition p is two times more credible than proposition q ”?

It is much safer to consider information of type (a), because information of type (b) is rather meaningless for a human.

Since fuzzy generalization of rough set theory using DRSA takes into account only *ordinal properties* of fuzzy membership degrees, it is the proper way of fuzzy generalization of rough set theory.

Moreover, the classical rough set approach (Pawlak 1982, 1991) can be seen as a specific case of our general model. This is important for several reasons. In particular, this interpretation of DRSA gives an insight into fundamental properties of the classical rough set approach and permits its further generalization.

Rough set theory (see Słowiński et al. (2014) for the state of the art) relies on the idea that some knowledge (data, information) is available about objects of a universe of discourse U . Thus, a subset of U is defined using the available knowledge about the objects and not on the base of information about membership or non-membership of the objects to the subset. For example, knowledge about patients suffering from a certain disease may contain information about body temperature, blood pressure, etc. All patients described by the same information are *indiscernible* in view of the available knowledge, and form groups of similar objects. These groups are called *elementary sets* and can be considered as basic granules of the available knowledge about patients. Elementary sets can be combined into compound concepts. For example, elementary sets of patients can be used to represent a set of patients suffering from a certain disease. Any union of elementary sets is called crisp set, while other sets are referred to as *rough sets*. Each rough set has boundary line objects, i.e., objects which, in view of the available knowledge, cannot be classified with certainty as members of the set or of its complement. Therefore, in the rough set approach, any set is associated with a pair of crisp sets, called the lower and the upper approximation. Intuitively, in view of the available information, the *lower approximation* consists of all objects which certainly belong to the set and the *upper approximation* contains all objects which possibly belong to the set. The difference between the upper and the lower approximation constitutes the *boundary* region of the rough set. Analogously, for a partition of universe U into classes, one may consider rough approximation of the partition. It appeared to be particularly useful for analysis of classification problems, being the most common decision problems.

The rough set approach operates on an *information table* composed of a set U of objects described by a set Q of attributes. If in the set Q disjoint sets (C and D) of *condition* and *decision* attributes are distinguished, then the information table is called *decision table*. It is often assumed, without loss of generality, that set D is a singleton $\{d\}$, and thus

decision attribute d makes a partition of set U into decision classes corresponding to its values. Data collected in such a decision table correspond to a multiple attribute classification problem. The classical indiscernibility-based rough set approach (IRSA) is naturally adapted to analysis of this type of decision problems, because the set of objects can be identified with examples of classification and it is possible to extract all the essential knowledge contained in the decision table using indiscernibility or similarity relations. However, as pointed out by the authors (see, e.g., Greco et al. (2001a, 2002a, 2004d) and Słowiński et al. (2002b)), IRSA cannot extract all the essential knowledge contained in the decision table if a background knowledge about *monotonic relationships* between evaluation of objects on condition attributes and their assignment to decision classes has to be taken into account. Such a background knowledge is typical for data describing various phenomena, as well as for data describing multiple-criteria decision problems (see, e.g., Figueira et al. (2005)), e.g., “the larger the mass and the smaller the distance, the larger the gravity,” “the more a tomato is red, the more it is ripe,” or “the better the school marks of a pupil, the better his overall classification.” The monotonic relationships, typical for multiple-criteria decision problems, follow from preferential ordering of value sets of attributes (scales of criteria), as well as preferential ordering of decision classes.

In order to take into account the *ordinal* properties of the considered attributes and the monotonic relationships between condition and decision attributes, a number of methodological changes to the original rough set theory were necessary. The main change was the replacement of the indiscernibility relation with a *dominance relation*, which permits approximation of ordered sets. The dominance relation is a very natural and rational concept within multiple-criteria decision analysis. The dominance-based rough set approach (DRSA) has been proposed and characterized by the authors (see, e.g., Greco et al. (2001a, 2002a, 2004b, c, d) and Słowiński et al. (2002b)).

Let us mention that ordered domains of attributes and a kind of order dependency among attributes has also been considered by specialists of relational databases (see, e.g., Ginsburg and Hull

(1983)). There is, however, a striking difference between consideration of orders in database queries and consideration of orders in knowledge discovery. More precisely, assuming ordered domains of attributes, knowledge discovery tends to discover monotonic relationships between ordered attributes, e.g., if a student is at least medium in networks and at least good in databases, then his overall evaluation is at least medium. On the other hand, assuming an order of attribute domains and order dependency in databases, one can exploit this given information for a more efficient answer to a query, e.g., when the dates of bank checks and their numbers are ordered, there is an order dependency between these two attributes because in day x a check cannot hold a number smaller than checks from day $x - 1$, which permits to prune the search tree and make the search more efficient.

Looking at DRSA from granular computing perspective, one can observe that DRSA permits to deal with *ordered data* by considering a specific type of information granules defined by means of *dominance-based constraints* having a syntax of the type: “ x is at least R ” or “ x is at most R ,” where R is a qualifier from a properly ordered scale. In evaluation space, such granules are *dominance cones*. In this sense, the contribution of DRSA consists in:

- Extending the paradigm of granular computing to problems involving ordered data
- Specifying a proper syntax and modality of information granules (the dominance-based constraints which should be adjoined to other modalities of information constraints, such as possibilistic, veristic and probabilistic (Zadeh 1999))
- Defining a methodology dealing properly with this type of information granules, and resulting in a theory of computing with words and reasoning about data in case of ordered data

Let us observe that other modalities of information constraints, such as veristic, possibilistic, and probabilistic, have also to deal with ordered values (with qualifiers relative to grades of truth, possibility, and probability). Therefore, that granular computing with ordered data and DRSA as a

proper way of reasoning about ordered data are very important in the future development of the whole domain of granular computing.

Indeed the DRSA approach proposed in Greco et al. (2000a, b) avoids arbitrary choice of fuzzy connectives and not meaningful operations on membership degrees. It exploits only ordinal character of the membership degrees and proposes a methodology of fuzzy rough approximation that infers the most cautious conclusion from available imprecise information. In particular, any approximation of knowledge about concept Y using knowledge about concept X is based on positive or negative relationships between premises and conclusions, i.e.:

- (i) “The more x is X , the more it is Y ” (positive relationship).
- (ii) “The more x is X , the less it is Y ” (negative relationship).

The following simple relationships illustrate (i) and (ii):

- “The larger the market share of a company, the greater its profit” (positive relationship).
- “The greater the debt of a company, the smaller its profit” (negative relationship).

These relationships have the form of *gradual decision rules* (Dubois and Prade 1992a). Examples of these decision rules are: “if a car is speedy with credibility at least 0.8 and it has high fuel consumption with credibility at most 0.7, then it is a good car with a credibility at least 0.9,” and “if a car is speedy with credibility at most 0.5 and it has high fuel consumption with credibility at least 0.8, then it is a good car with a credibility at most 0.6.” Remark that the syntax of gradual decision rules is based on monotonic relationships between degrees of credibility that can also be found in dominance-based decision rules induced from preference-ordered data. This explains why one can build a fuzzy rough approximation using DRSA.

Finally, the fuzzy rough approximation taking into account monotonic relationships can be

applied to case-based reasoning (Greco et al. 2006a). In this perspective, it is interesting to consider monotonicity of the type “the more similar is y to x , the more credible is that y belongs to the same class as x .” Application of DRSA in this context leads to decision rules similar to the gradual decision rules:

the more object z is similar to a referent object x w.r.
t. condition attribute s , the more z is similar to a
referent object x w.r.t. decision attribute t ,

or, equivalently, but more technically,

$$s(z, x) \geq \alpha \Rightarrow t(z, x) \geq \alpha$$

where functions s and t measure the credibility of similarity with respect to condition attribute and decision attribute, respectively. When there are multiple condition and decision attributes, functions s and t aggregate similarity with respect to these attributes.

The decision rules induced on the basis of DRSA do not need the aggregation of the similarity with respect to different attributes into one comprehensive similarity. This is important because it permits to avoid using aggregation operators (weighted average, min, etc.) which are always arbitrary to some extent. Moreover, the DRSA decision rules permit to consider different thresholds for degrees of credibility in the premise and in the conclusion.

The article is organized as follows. Section “Glossary” presents the philosophical basis of DRSA granular computing. Section “Definition of the Subject and Its Importance” recalls the main steps of DRSA for multiple-criteria classification, or, more in general, for ordinal classification. Section “Introduction: Granular Computing and Ordered Data” presents, a review of fuzzy set extensions of DRSA based on fuzzy connectives. In section “Philosophical Basis of DRSA Granular Computing,” a “probabilistic” version of DRSA, the variable-consistency dominance-based rough set approach is presented. Dominance-based rough approximation of a fuzzy set is presented in section “Dominance-Based Rough Set Approach.” The explanation that IRSA is a particular case of DRSA is presented in section “Fuzzy Set

Extensions of the Dominance-Based Rough Set Approach.” Section “Variable-Consistency Dominance-Based Rough Set Approach (VC-DRSA)” is devoted to DRSA for case-based reasoning. In section “Dominance-Based Rough Approximation of a Fuzzy Set,” an algebra modeling logic of DRSA is presented. Section “Monotonic Rough Approximation of a Fuzzy Set Versus Classical Rough Set” contains conclusions.

Philosophical Basis of DRSA Granular Computing

It is interesting to analyze the relationships between DRSA and granular computing from the point of view of the philosophical basis of rough set theory proposed by Pawlak. Since according to Pawlak (2001), rough set theory refers to some ideas of Gottlob Frege (vague concepts), Gottfried Leibniz (indiscernibility), George Boole (reasoning methods), Jan Łukasiewicz (multi-valued logic), and Thomas Bayes (inductive reasoning), it is meaningful to give an account for DRSA generalization of rough sets, justifying it in reference to some of these main ideas recalled by Pawlak.

The *identity of indiscernibles* is a principle of analytic ontology first explicitly formulated by Gottfried Leibniz in his *Discourse on Metaphysics*, (Loemker 1969). Two objects x and y are defined indiscernible, if x and y have the same properties. The principle of identity of indiscernibles states that

if x and y are indiscernible, then $x = y$.
(II1)

This can be expressed also as *if $x \neq y$, then x and y are discernible*, i.e., there is at least one property that x has and y does not, or vice versa. The converse of the principle of identity of indiscernibles is called *indiscernibility of identicals* and states that *if $x = y$, then x and y are indiscernible*, i.e., they have the same properties. This is equivalent to say that if there is at least one property that x has and y does not, or vice versa, then $x \neq y$. The conjunction of both principles is often referred to as “Leibniz’s law.”

Rough set theory is based on a weaker interpretation of Leibniz’s law, having as objective the ability to classify objects falling under the same concept. This reinterpretation of the Leibniz’s law is based on a reformulation of the principle of identity of indiscernibles as follows:

if x and y are indiscernible, then x and y belong to the same class. (II2)

Let us observe that the word “class” in the previous sentence can be considered as synonymous of “granule.” Thus, from the point of view of granular computing, II2 can be rewritten as

*if x and y are indiscernible,
then x and y belong to the same granule of
classification.* (II2')

Notice also that the principle of indiscernibility of identicals cannot be reformulated in analogous terms. In fact, such an analogous reformulation would amount to state that *if x and y belong to the same class, then x and y are indiscernible*. This principle is too strict, however, because there can be two discernible objects x and y belonging to the same class. Thus, within rough set theory, the principle of indiscernibility of identicals should continue to hold in its original formulation (i.e., *if $x = y$, then x and y are indiscernible*). It is worthwhile to observe that the relaxation in the consequence of the implication from II1 to II2 implies an implicit relaxation also in the antecedent. In fact, one could say that two objects are identicals if they have the same properties, if one would be able to take into account all conceivable properties. For human limitations this is not the case, therefore, one can imagine that II2' can be properly reformulated as

*if x and y are indiscernible taking into
account a given set of properties,
then x and y belong to the same class.* (II2'')

This weakening in the antecedent of the implication means also that the objects indiscernible with respect to a given set of properties can be seen as a granule, such that, finally, the II2 could be rewritten in terms of granulation as

if x and y belong to the same granule with respect to a given set of properties, then x and y belong to the same classification granule. (II2'')

For this reason, rough set theory needs a still weaker form of the principle of identity of indiscernibles. Such a principle can be formulated using the idea of vagueness due to Gottlob Frege. According to Frege “the concept must have a sharp boundary – to the concept without a sharp boundary there would correspond an area that had not a sharp boundary-line all around.” Therefore, following this intuition, the principle of identity of indiscernibles can be further reformulated as

if x and y are indiscernible, then x and y should belong to the same class. (II3)

In terms of granular computing, II3 can be rewritten as

if x and y belong to the same granule with respect to a given set of properties, then x and y should belong to the same classification granule. (II3')

This reformulation of the principle of identity of indiscernibles implies that there is an inconsistency in the statement that x and y are indiscernible, and x and y belong to different classes. Thus, the Leibniz's principle of identity of indiscernibles and the Frege's intuition about vagueness found the basic idea of the rough set concept proposed by Pawlak.

The above reconstruction of the basic idea of the Pawlak's rough set should be completed, however, by referring to another basic idea. This is the idea of Georg Boole that concerns a property which is satisfied or not satisfied. It is quite natural to weaken this principle admitting that a property can be satisfied to some degree. This idea of graduality can be attributed to Jan Łukasiewicz and his proposal of many-valued logic where, in addition to well-known truth values “true” and “false,” other truth values representing partial degrees of truth were present. The Łukasiewicz's idea of graduality has been reconsidered,

generalized and fully exploited by Zadeh (1965) within fuzzy set theory, where graduality concerns membership to a set.

In this sense, any proposal of putting rough sets and fuzzy sets together can be seen as a reconstruction of the rough set concept, where the Boole's idea of binary logic is abandoned in favor of the Łukasiewicz's idea of many-valued logic, such that the Leibniz's principle of identity of indiscernibles and the Frege's intuition about vagueness are combined with the idea that a property is satisfied to some degree.

Putting aside, for the moment, the Frege's intuition about vagueness, but taking into account the concept of graduality, the principle of identity of indiscernibles can be reformulated as follows:

if the grade of each property for x is greater than or equal to the grade for y , then x belongs to the considered class in a grade at least as high as y . (II4)

Taking into account the paradigm of granular computing, II4 can be rewritten as

if x belongs to the granules defined by considered properties more than y , because the grade of each property for x is greater than or equal to the grade for y , then x belongs to the considered classification granule in a grade at least as high as y . (II4')

Considering the concept of graduality together with the Frege's intuition about vagueness, one can reformulate the principle of identity of indiscernibles as follows:

if the grade of each property for x is greater than or equal to the grade for y then x should belong to the considered class in a grade at least as high as y . (II5)

In terms of granular computing, II5 can be rewritten as

if x belongs to the granules defined by considered properties

more than y , because the grade of each property for x is greater than or equal to the grade for y , then x should belong to the considered classification granule in a grade at least as high as y . (II5')

The formulation II5' of the principle of identity of indiscernibles is perfectly concordant with the rough set concept defined within the dominance-based rough set approach (DRSA) (Greco et al. 2002a).

DRSA has been proposed by the authors to deal with ordinal properties of data related to preferences in decision problems (Greco et al. 2004d; Słowiński et al. 2002b). The fundamental feature of DRSA is that it handles monotonicity of comprehensive evaluation of objects with respect to preferences relative to evaluation of these objects on particular attributes. For example, the more preferred is a car with respect to such attributes as maximum speed, acceleration, fuel consumption, and price, the better is its comprehensive evaluation. The type of monotonicity considered within DRSA is also meaningful for problems where relationships between different aspects of a phenomenon described by data are to be taken into account, even if preferences are not considered. Indeed, monotonicity concerns, in general, mutual trends existing between different variables, like distance and gravity in physics, or inflation rate and interest rate in economics. Whenever a relationship between different aspects of a phenomenon is discovered, this relationship can be represented by a monotonicity with respect to some specific measures of the considered aspects. Formulation II5 of the principle of identity of indiscernibles refers to this type of monotonic relationships. So, in general, the monotonicity permits to translate into a formal language a primitive intuition of relationships between different concepts of our knowledge corresponding to the principle of identity of indiscernibles formulated as II5'.

Dominance-Based Rough Set Approach

This section presents the main concepts of the dominance-based rough set approach (DRSA) (for a more complete presentation see, e.g.,

Greco et al. (2001a, 2002a, 2004d) and Słowiński et al. (2002b, 2014).

Information about objects is represented in the form of an information table. The rows of the table are labeled by objects, whereas columns are labeled by attributes and entries of the table are attribute values. Formally, an information system (table) is the 4-tuple $S = \langle U, Q, V, \phi \rangle$, where U is a finite set of objects, Q is a finite set of attributes, $V = \bigcup_{q \in Q} V_q$ and V_q is the set of values of the attribute q , and $\phi : U \times Q \rightarrow V_q$ is a total function such that $\phi(x, q) \in V_q$ for every $q \in Q$, $x \in U$, called an information function (Pawlak 1991). The set Q is, in general, divided into set C of condition attributes and set D of decision attributes.

Condition attributes with values sets ordered according to decreasing or increasing preference are called *criteria*. For criterion $q \in Q$, $\succsim_q$ is a *weak preference* relation on U such that $x \succsim_q y$ means “ x is at least as good as y with respect to criterion q .” It is supposed that $\succsim_q$ is a complete preorder, i.e., a strongly complete and transitive binary relation, defined on U on the basis of evaluations $\phi(\cdot, q)$. Without loss of generality, the preference is supposed to increase with the value of $\phi(\cdot, q)$ for every criterion $q \in C$, such that for all $x, y \in U$, $x \succsim_q y$ if and only if $\phi(x, q) \geq \phi(y, q)$.

Furthermore, it is supposed that the set of decision attributes D is a singleton $\{d\}$. Values of decision attribute d makes a partition of U into a finite number of decision classes, $\mathbf{CI} = \{Cl_t, t = 1, \dots, n\}$, such that each $x \in U$ belongs to one and only one class $Cl_t \in \mathbf{CI}$. It is supposed that the classes are preference ordered, i.e., for all $r, s \in \{1, \dots, n\}$, such that $r > s$, the objects from Cl_r are preferred to the objects from Cl_s . More formally, if $\succsim$ is a *comprehensive weak preference relation* on U , i.e., if for all $x, y \in U$, $x \succsim y$ means “ x is at least as good as y ”; it is supposed: $[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow [x \succsim y \text{ and not } y \succsim x]$. The above assumptions are typical for consideration of *ordinal classification problems* (also called *multiple-criteria sorting problems*).

The sets to be approximated are called *upward union* and *downward union* of classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \quad t = 1, \dots, n.$$

The statement $x \in Cl_t^{\geq}$ means “ x belongs to at least class Cl_t ,” while $x \in Cl_t^{\leq}$ means “ x belongs to at most class Cl_t .” Let us remark that $Cl_1^{\geq} = Cl_n^{\leq} = U$, $Cl_n^{\geq} = Cl_1^{\leq}$, and $Cl_1^{\geq} = Cl_1$. Furthermore, for $t = 2, \dots, n$,

$$Cl_{t-1}^{\leq} = U - Cl_t^{\geq} \quad \text{and} \quad Cl_t^{\geq} = U - Cl_{t-1}^{\leq}.$$

The key idea of the rough set approach is representation (approximation) of knowledge generated by decision attributes, by “*granules of knowledge*” generated by condition attributes.

In DRSA, where condition attributes are criteria and decision classes are preference ordered, the represented knowledge is a collection of *upward* and *downward unions of classes*, and the “*granules of knowledge*” are sets of objects defined using a dominance relation.

x dominates y with respect to $P \subseteq C$ (shortly, x P -dominates y), denoted by $x D_P y$, if for every criterion $q \in P$, $\phi(x, q) \geq \phi(y, q)$. The relation of P -dominance is reflexive and transitive, that is, a partial preorder.

Given a set of criteria $P \subseteq C$ and $x \in U$, the “*granules of knowledge*” used for approximation in DRSA are:

- A set of objects dominating x , called *P-dominating set*,
 $D_P^+(x) = \{y \in U: y D_P x\}.$
- A set of objects dominated by x , called *P-dominated set*,
 $D_P^-(x) = \{y \in U: x D_P y\}.$

Remark that the “*granules of knowledge*” defined above has the form of upward (positive) and downward (negative) *dominance cones* in the evaluation space.

Let us recall that the *dominance principle* (or Pareto principle) requires that an object x dominating object y on all considered criteria (i.e., x having evaluations at least as good as y on all considered criteria) should also dominate y on the decision (i.e., x should be assigned to at least

as good decision class as y). This principle is the only objective principle that is widely agreed upon in the multiple-criteria comparisons of objects.

Given $P \subseteq C$, the inclusion of an object $x \in U$ to the upward union of classes $Cl_t^{\geq}$, $t = 2, \dots, n$, is *inconsistent with the dominance principle* if one of the following conditions holds:

- x belongs to class Cl_t or better, but it is P -dominated by an object y belonging to a class worse than Cl_t , i.e., $x \in Cl_t^{\geq}$ but $D_P^+(x) \cap Cl_{t-1}^{\leq} \neq \emptyset$.
- x belongs to a worse class than Cl_t , but it P -dominates an object y belonging to class Cl_t or better, i.e., $x \notin Cl_t^{\geq}$ but $D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset$.

If, given a set of criteria $P \subseteq C$, the inclusion of $x \in U$ to $Cl_t^{\geq}$, where $t=2, \dots, n$, is inconsistent with the dominance principle, then x belongs to $Cl_t^{\geq}$ with some ambiguity. Thus, x belongs to $Cl_t^{\geq}$ without any ambiguity with respect to $P \subseteq C$, if $x \in Cl_t^{\geq}$ and there is no inconsistency with the dominance principle. This means that all objects P -dominating x belong to $Cl_t^{\geq}$, i.e., $D_P^+(x) \subseteq Cl_t^{\geq}$.

Furthermore, x possibly belongs to $Cl_t^{\geq}$ with respect to $P \subseteq C$, if one of the following conditions holds:

- According to decision attribute d , x belongs to $Cl_t^{\geq}$.
- According to decision attribute d , x does not belong to $Cl_t^{\geq}$, but it is inconsistent in the sense of the dominance principle with an object y belonging to $Cl_t^{\geq}$.

In terms of ambiguity, x possibly belongs to $Cl_t^{\geq}$ with respect to $P \subseteq C$, if x belongs to $Cl_t^{\geq}$ with or without any ambiguity. Due to the reflexivity of the dominance relation D_P , the above conditions can be summarized as follows: x possibly belongs to class Cl_t or better, with respect to $P \subseteq C$, if among the objects P -dominated by x there is an object y belonging to class Cl_t or better, i.e., $D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset$.

The P -lower approximation of $Cl_t^{\geq}$, denoted by $\underline{P}Cl_t^{\geq}$, and the P -upper approximation of $Cl_t^{\geq}$, denoted by $\overline{P}(Cl_t^{\geq})$, are defined as follows ($t = 1, \dots, n$):

$$\begin{aligned}\underline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\}, \\ \overline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}.\end{aligned}$$

Analogously, one can define the P -lower approximation and the P -upper approximation of $Cl_t^{\leq}$ as follows ($t = 1, \dots, n$):

$$\begin{aligned}\underline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\}, \\ \overline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}.\end{aligned}$$

The P -lower and P -upper approximations so defined satisfy the following inclusion properties for each $t \in \{1, \dots, n\}$ and for all $P \subseteq C$:

$$\begin{aligned}\underline{P}(Cl_t^{\geq}) &\subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \\ \underline{P}(Cl_t^{\leq}) &\subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}).\end{aligned}$$

The P -lower and P -upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ have an important complementarity property, according to which,

$$\begin{aligned}\underline{P}(Cl_t^{\geq}) &= U - \overline{P}(Cl_{t-1}^{\leq}) \text{ and} \\ \overline{P}(Cl_t^{\geq}) &= U - \underline{P}(Cl_{t-1}^{\leq}), t = 2, \dots, n,\end{aligned}$$

$$\begin{aligned}\underline{P}(Cl_t^{\leq}) &= U - \overline{P}(Cl_{t+1}^{\geq}) \text{ and} \\ \overline{P}(Cl_t^{\leq}) &= U - \underline{P}(Cl_{t+1}^{\geq}), t = 1, \dots, n-1.\end{aligned}$$

The P -boundary of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$ respectively, are defined as follows ($t = 1, \dots, n$):

$$\begin{aligned}Bn_P(Cl_t^{\geq}) &= \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \\ Bn_P(Cl_t^{\leq}) &= \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}).\end{aligned}$$

Due to complementarity property, $Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq})$, for $t = 2, \dots, n$.

The dominance-based rough approximations of upward and downward unions of classes can

serve to induce “if..., then...” decision rules. It is meaningful to consider the following five types of decision rules:

- (1) Certain $D_{\geq}$ -decision rules: if $x_{q1} \succ_{q1} r_{q1}$ and $x_{q2} \succ_{q2} r_{q2}$ and ... $x_{qp} \succ_{qp} r_{qp}$, then certainly x belongs to $Cl_t^{\geq}$, where, for each $w_q, z_q \in X_q$, “ $w_q \succ_{qz_q}$ ” means “ w_q is at least as good as z_q .”
- (2) Possible $D_{\geq}$ -decision rules: if $x_{q1} \succ_{q1} r_{q1}$ and $x_{q2} \succ_{q2} r_{q2}$ and ... $x_{qp} \succ_{qp} r_{qp}$, then x possibly belongs to $Cl_t^{\geq}$.
- (3) Certain $D_{\leq}$ -decision rules: if $x_{q1} \preccurlyeq_{q1} r_{q1}$ and $x_{q2} \preccurlyeq_{q2} r_{q2}$ and ... $x_{qp} \preccurlyeq_{qp} r_{qp}$, then certainly x belongs to $Cl_t^{\leq}$, where, for each $w_q, z_q \in X_q$, “ $w_q \preccurlyeq_{qz_q}$ ” means “ w_q is at most as good as z_q .”
- (4) Possible $D_{\leq}$ -decision rules: if $x_{q1} \preccurlyeq_{q1} r_{q1}$ and $x_{q2} \preccurlyeq_{q2} r_{q2}$ and ... $x_{qp} \preccurlyeq_{qp} r_{qp}$, then x possibly belongs to $Cl_t^{\leq}$.
- (5) Approximate $D_{\geq \leq}$ -decision rules: if $x_{q1} \succ_{q1} r_{q1}$ and ... $x_{qk} \succ_{qk} r_{qk}$ and $x_{q(k+1)} \preccurlyeq_{q(k+1)} r_{q(k+1)}$ and ... $x_{qp} \preccurlyeq_{qp} r_{qp}$, then $x \in Cl_s^{\geq} \cap Cl_t^{\leq}$, where $s < t$.

The rules of type (1) and (3) represent certain knowledge extracted from the decision table, while the rules of type (2) and (4) represent possible knowledge. Rules of type (5) represent doubtful knowledge.

Fuzzy Set Extensions of the Dominance-Based Rough Set Approach

The concept of dominance can be refined by introducing gradedness through the use of fuzzy sets. Here are basic definitions of fuzzy connectives (Fodor and Roubens 1994; Klement et al. 2000). For each proposition p , one can consider its truth value $v(p)$ ranging from $v(p) = 0$ (p is definitely false) to $v(p) = 1$ (p is definitely true); and for all intermediate values, the greater $v(p)$, the more credible is the truth of p . A negation is a non-increasing function $N : [0, 1] \rightarrow [0, 1]$ such that $N(0) = 1$ and $N(1) = 0$. Given proposition p , $N(v(p))$ states the credibility of the negation of p . A t-norm T and a t-conorm T^* are two functions $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ and $T^* : [0, 1] \times [0, 1] \rightarrow [0, 1]$, such that

given two propositions, p and q , $T(v(p), v(q))$ represents the credibility of the conjunction of p and q , and $T^*(v(p), v(q))$ represents the credibility of the

disjunction of p and q . t-norm T and t-conorm T^* must satisfy the following properties:

$$\begin{aligned} T(\alpha, \beta) &= T(\beta, \alpha) \text{ and } T^*(\alpha, \beta) = T^*(\beta, \alpha), \text{ for all } \alpha, \beta \in [0, 1], \\ T(\alpha, \beta) &\leq T(\gamma, \delta) \text{ and } T^*(\alpha, \beta) \leq T^*(\gamma, \delta), \text{ for all } \alpha, \beta, \gamma, \delta \in [0, 1] \\ &\text{such that } \alpha \leq \gamma \text{ and } \beta \leq \delta, \\ T(\alpha, T(\beta, \gamma)) &= T(T(\alpha, \beta), \gamma) \text{ and } T^*(\alpha, T^*(\beta, \gamma)) = T^*(T^*(\alpha, \beta), \gamma), \\ &\text{for all } \alpha, \beta, \gamma \in [0, 1], \\ T(1, \alpha) &= \alpha \text{ and } T^*(0, \alpha) = \alpha, \text{ for all } \alpha \in [0, 1]. \end{aligned}$$

A negation is strict if it is strictly decreasing and continuous. A negation N is involutive if, for all $\alpha \in [0, 1]$, $N(N(\alpha)) = \alpha$. A strong negation is an involutive strict negation. If N is a strong negation, then (T, T^*, N) is a de Morgan triplet if $N(T^*(\alpha, \beta)) = T(N(\alpha), N(\beta))$. A fuzzy implication is a function $I: [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that, given two propositions p and q , $I(v(p), v(q))$ represents the credibility of the implication of q by p . A fuzzy implication must satisfy the following properties (see Fodor and Roubens (1994)):

$$\begin{aligned} I(\alpha, \beta) &\geq I(\gamma, \beta) \text{ for all } \alpha, \beta, \gamma \in [0, 1], \text{ such that } \alpha \leq \gamma, \\ I(\alpha, \beta) &\geq I(\alpha, \gamma) \text{ for all } \alpha, \beta, \gamma \in [0, 1], \text{ such that } \beta \geq \gamma, \\ I(0, \alpha) &= 1, I(\alpha, 1) = 1 \text{ for all } \alpha \in [0, 1], \\ I(1, 0) &= 0. \end{aligned}$$

An implication I_{N, T^*}^- is a T^* -implication if there is a t-conorm T^* and a strong negation N such that $I_{N, T^*}^-(\alpha, \beta) = T^*(N(\alpha), \beta)$. A fuzzy similarity relation on the universe U is a fuzzy binary relation (i.e., function $R: U \times U \rightarrow [0, 1]$), reflexive ($R(x, x) = 1$ for all $x \in U$), symmetric ($R(x, y) = R(y, x)$ for all $x, y \in U$), and transitive (given t-norm T , $T(R(x, y), R(y, z)) \leq R(x, z)$ for all $x, y, z \in U$).

Let $\succsim_q$ be a fuzzy weak preference relation on U with respect to criterion $q \in C$, i.e., $\succsim_q: U \times U \rightarrow [0, 1]$, such that, for all $x, y \in U$, $\succsim_q(x, y)$ represents the credibility of the proposition “ x is at least as good as y with respect to criterion q .” Suppose that $\succsim_q$ is a fuzzy partial T -preorder, i.e., that it is reflexive ($\succsim_q(x, x) = 1$ for each $x \in U$) and T -transitive (T

($\succsim_q(x, y), \succsim_q(y, z)) \leq \succsim_q(x, z)$, for each $x, y, z \in U$) (see Fodor and Roubens (1994)). Using the fuzzy weak preference relations $\succsim_q, q \in C$, a fuzzy dominance relation on U (denotation $D_P(x, y)$) can be defined, for all $P \subseteq C$, as follows:

$$D_P(x, y) = T_{q \in P}(\succsim_q(x, y)).$$

Given $(x, y) \in U \times U$, $D_P(x, y)$ represents the credibility of the proposition “ x is at least as good as y with respect to each criterion q from P .” Since the fuzzy weak preference relations $\succsim_q$ are supposed to be partial T -preorders, then also the fuzzy dominance relation D_P is a partial T -preorder. Furthermore, let $Cl = \{Cl_t, t = 1, \dots, n\}$ be a set of fuzzy classes in U such that, for each $x \in U$, $Cl_t(x)$ represents the membership function of x to Cl_t . It is supposed, as before, that the classes of Cl are increasingly ordered, i.e., that for all $r, s \in \{1, \dots, n\}$ such that $r > s$, the objects from Cl_r have a better comprehensive evaluation than the objects from Cl_s . On the basis of the membership functions of the fuzzy class Cl_t , fuzzy membership functions of two other sets can be defined as follows:

- (1) The upward union fuzzy set $Cl_t^{\geq}$, whose membership function $Cl_t^{\geq}(x)$ represents the credibility of the proposition “ x is at least as good as the objects in Cl_t ”:

$$\begin{aligned} Cl_t^{\geq}(x) &= \begin{cases} 1 & \text{if } \exists s \in \{1, \dots, n\} : Cl_s(x) > 0 \text{ and } s > t \\ Cl_t(x) & \text{otherwise} \end{cases}; \end{aligned}$$

- (2) The downward union fuzzy set $Cl_t^{\leq}$, whose membership function $Cl_t^{\leq}(x)$ represents the credibility of the proposition “ x is at most as good as the objects in Cl_t ”:

$$Cl_t^{\leq}(x) = \begin{cases} 1 & \text{if } \exists s \in \{1, \dots, n\} : Cl_s(x) > 0 \text{ and } s < t \\ Cl_t(x) & \text{otherwise} \end{cases}.$$

The P -lower and the P -upper approximations of $Cl_t^{\geq}$ with respect to $P \subseteq C$ are fuzzy sets in U , whose membership functions, denoted by $\underline{P}[Cl_t^{\geq}(x)]$ and $\overline{P}[Cl_t^{\geq}(x)]$, respectively, are defined as:

$$\begin{aligned} \underline{P}[Cl_t^{\geq}(x)] &= T_{y \in U}(T^*(N(D_P(y, x)), Cl_t^{\geq}(y))), \\ \overline{P}[Cl_t^{\geq}(x)] &= T_{y \in U}^*(T(D_P(x, y), Cl_t^{\geq}(y))). \end{aligned}$$

$\underline{P}[Cl_t^{\geq}(x)]$ represents the credibility of the proposition “for all $y \in U$, y does not dominate x with respect to criteria from P or y belongs to $Cl_t^{\geq}$,” while $\overline{P}[Cl_t^{\geq}(x)]$ represents the credibility of the proposition “there is at least one $y \in U$ dominated by x with respect to criteria from P which belongs to $Cl_t^{\geq}$.”

The P -lower and P -upper approximations of $Cl_t^{\leq}$ with respect to $P \subseteq C$, denoted by $\underline{P}[Cl_t^{\leq}(x)]$ and $\overline{P}[Cl_t^{\leq}(x)]$, respectively, can be defined, analogously, as:

$$\begin{aligned} \underline{P}[Cl_t^{\leq}(x)] &= T_{y \in U}(T^*(N(D_P(x, y)), Cl_t^{\leq}(y))), \\ \overline{P}[Cl_t^{\leq}(x)] &= T_{y \in U}^*(T(D_P(y, x), Cl_t^{\leq}(y))). \end{aligned}$$

$\underline{P}[Cl_t^{\leq}(x)]$ represents the credibility of the proposition “for all $y \in U$, x does not dominate y with respect to criteria from P or y belongs to $Cl_t^{\leq}$,” while $\overline{P}[Cl_t^{\leq}(x)]$ represents the credibility of the proposition “there is at least one $y \in U$ dominating x with respect to criteria from P which belongs to $Cl_t^{\leq}$.”

Let us remark that, using the definition of the T^* -implication, it is possible to rewrite the definitions of $\underline{P}[Cl_t^{\geq}(x)]$, $\overline{P}[Cl_t^{\geq}(x)]$, $\underline{P}[Cl_t^{\leq}(x)]$, and $\overline{P}[Cl_t^{\leq}(x)]$, in the following way:

$$\begin{aligned} \underline{P}[Cl_t^{\geq}(x)] &= T_{y \in U}(I_{T^*, N}^-(D_P(y, x), Cl_t^{\geq}(y))), \\ \overline{P}[Cl_t^{\geq}(x)] &= T_{y \in U}^*(N(I_{T^*, N}^-(D_P(x, y), N(Cl_t^{\geq}(y))))), \\ \underline{P}[Cl_t^{\leq}(x)] &= T_{y \in U}(I_{T^*, N}^-(D_P(x, y), Cl_t^{\leq}(y))), \\ \overline{P}[Cl_t^{\leq}(x)] &= T_{y \in U}^*(N(I_{T^*, N}^-(D_P(y, x), N(Cl_t^{\leq}(y))))). \end{aligned}$$

The following results can be proved:

- (1) For each $x \in U$ and for each $t \in \{1, \dots, n\}$,

$$\begin{aligned} \underline{P}[Cl_t^{\geq}(x)] &\leq Cl_t^{\geq}(x) \leq \overline{P}[Cl_t^{\geq}(x)], \quad \underline{P}[Cl_t^{\leq}(x)] \\ &\leq Cl_t^{\leq}(x) \leq \overline{P}[Cl_t^{\leq}(x)]; \end{aligned}$$

- (2) If (T, T^*, N) constitute a de Morgan triplet and if $N[Cl_t^{\geq}(x)] = Cl_{t-1}^{\leq}(x)$ for each $x \in U$ and $t = 2, \dots, n$, then

$$\begin{aligned} \underline{P}[Cl_t^{\geq}(x)] &= N(\overline{P}[Cl_{t-1}^{\leq}(x)]), \quad \overline{P}[Cl_t^{\geq}(x)] \\ &= N(\underline{P}[Cl_{t-1}^{\leq}(x)]), \quad t = 2, \dots, n, \end{aligned}$$

$$\begin{aligned} \underline{P}[Cl_t^{\leq}(x)] &= N(\overline{P}[Cl_{t+1}^{\geq}(x)]), \quad \overline{P}[Cl_t^{\leq}(x)] \\ &= N(\underline{P}[Cl_{t+1}^{\geq}(x)]), \quad t = 1, \dots, n-1; \end{aligned}$$

- (3) For all $P \subseteq R \subseteq C$, for all $x \in U$, and for each $t \in \{1, \dots, n\}$,

$$\begin{aligned} \underline{P}[Cl_t^{\geq}(x)] &\leq \underline{R}[Cl_t^{\geq}(x)], \quad \overline{P}[Cl_t^{\geq}(x)] \\ &\geq \overline{R}[Cl_t^{\geq}(x)], \end{aligned}$$

$$\begin{aligned} \underline{P}[Cl_t^{\leq}(x)] &\leq \underline{R}[Cl_t^{\leq}(x)], \quad \overline{P}[Cl_t^{\leq}(x)] \\ &\geq \overline{R}[Cl_t^{\leq}(x)]. \end{aligned}$$

Results (1) to (3) can be read as fuzzy counterparts of the following results well-known within the classical rough set approach: (1) (*inclusion property*) says that $Cl_t^{\geq}$ and $Cl_t^{\leq}$ include their P -lower approximations and are included in their P -upper approximations; (2) (*complementarity property*) says that the P -lower (P -upper)

approximation of $Cl_t^{\geq}$ is the complement of the P -upper (P -lower) approximation of its complementary set $Cl_{t-1}^{\leq}$, (analogous property holds for $Cl_t^{\leq}$ and $Cl_{t+1}^{\geq}$); and (3) (*monotonicity with respect to sets of attributes*) says that enlarging the set of criteria, the membership to the lower approximation does not decrease and the membership to the upper approximation does not increase.

Greco, Inuiguchi, and Słowiński (Greco et al. 1999) proposed, moreover, the following fuzzy rough approximations based on dominance, which go in line with the fuzzy rough approximation by Dubois and Prade (1990, 1992b), concerning classical rough sets:

$$\underline{P}[Cl_t^{\geq}(x)] = \inf_{y \in U} (I(D_P(y, x), Cl_t^{\geq}(y))),$$

$$\overline{P}[Cl_t^{\geq}(x)] = \sup_{y \in U} (T(D_P(x, y), Cl_t^{\geq}(y))),$$

$$\underline{P}[Cl_t^{\leq}(x)] = \inf_{y \in U} (I(D_P(x, y), Cl_t^{\leq}(y))),$$

$$\overline{P}[Cl_t^{\leq}(x)] = \sup_{y \in U} (T(D_P(y, x), Cl_t^{\leq}(y))).$$

Using fuzzy rough approximations based on DRSA, one can induce decision rules having the same syntax as the decision rules obtained from crisp DRSA. In this case, however, each decision rule has a fuzzy credibility.

Variable-Consistency Dominance-Based Rough Set Approach (VC-DRSA)

The definitions of rough approximations introduced in section “[Definition of the Subject and Its Importance](#)” are based on a strict application of the dominance principle. However, when defining non-ambiguous objects, it is reasonable to accept a limited proportion of negative examples, particularly for large data tables. Such extended version of DRSA is called Variable-Consistency DRSA model (VC-DRSA) (Greco et al. 2016; Błaszczyński et al. 2009).

For any $P \subseteq C$, $x \in U$ belongs to $Cl_t^{\geq}$ *without any ambiguity* at consistency level $l \in (0, 1]$, if $x \in Cl_t^{\geq}$ and at least $l \cdot 100\%$ of all objects $y \in U$ dominating x with respect to P also belong to $Cl_t^{\geq}$, i.e., for $t = 2, \dots, n$,

$$\frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq l.$$

The level l is called *consistency level* because it controls the degree of consistency with respect to objects qualified as belonging to $Cl_t^{\geq}$ without any ambiguity. In other words, if $l < 1$, then at most $(1 - l) \cdot 100\%$ of all objects $y \in U$ dominating x with respect to P do not belong to $Cl_t^{\geq}$ and thus contradict the inclusion of x in $Cl_t^{\geq}$.

Analogously, for any $P \subseteq C$, $x \in U$ belongs to $Cl_t^{\leq}$ *without any ambiguity* at consistency level $l \in (0, 1]$, if $x \in Cl_t^{\leq}$ and at least $l \cdot 100\%$ of all the objects $y \in U$ dominated by x with respect to P also belong to $Cl_t^{\leq}$, i.e., for $t = 1, \dots, n - 1$,

$$\frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq l.$$

The concept of non-ambiguous objects at some consistency level l leads naturally to the corresponding definition of P -lower approximations of the unions of classes $Cl_t^{\geq}$ and $Cl_t^{\leq}$, respectively:

$$\underline{P}^l(Cl_t^{\geq}) = \left\{ x \in Cl_t^{\geq} : \frac{|D_P^+(x) \cap Cl_t^{\geq}|}{|D_P^+(x)|} \geq l \right\}, \quad t = 2, \dots, n,$$

$$\underline{P}^l(Cl_t^{\leq}) = \left\{ x \in Cl_t^{\leq} : \frac{|D_P^-(x) \cap Cl_t^{\leq}|}{|D_P^-(x)|} \geq l \right\}, \quad t = 1, \dots, n - 1.$$

Given $P \subseteq C$ and consistency level l , the corresponding P -upper approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $\overline{P}^l(Cl_t^{\geq})$ and $\overline{P}^l(Cl_t^{\leq})$, respectively, can be defined as a complement of $\underline{P}^l(Cl_{t-1}^{\leq})$ and $\underline{P}^l(Cl_{t+1}^{\geq})$ with respect to U :

$$\overline{P}^l(Cl_t^{\geq}) = U - \underline{P}^l(Cl_{t-1}^{\leq}), \quad t = 2, \dots, n,$$

$$\overline{P}^l(Cl_t^{\leq}) = U - \underline{P}^l(Cl_{t+1}^{\geq}), \quad t = 1, \dots, n - 1.$$

$\overline{P}^l(Cl_t^{\geq})$ can be interpreted as a set of all the objects belonging to $Cl_t^{\geq}$, *possibly ambiguous* at consistency level l . Analogously, $\overline{P}^l(Cl_t^{\leq})$ can be interpreted as a set of all the objects belonging to $Cl_t^{\leq}$, *possibly ambiguous* at consistency level l . The P -boundaries (P -doubtful regions) of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ at consistency level l are defined as:

$$\begin{aligned} Bn_P^l(CI_t^{\geq}) &= \bar{P}^l(CI_t^{\geq}) - \underline{P}^l(CI_t^{\geq}), \quad t=2, \dots, n \\ Bn_P^l(CI_t^{\leq}) &= \bar{P}^l(CI_t^{\leq}) - \underline{P}^l(CI_t^{\leq}), \quad t=1, \dots, n-1. \end{aligned}$$

The *variable-consistency* model of the dominance-based rough set approach provides some degree of flexibility in assigning objects to lower and upper approximations of the unions of decision classes. The following properties can be easily proved: for $0 < l' < l \leq 1$,

$$\begin{aligned} \underline{P}^l(CI_t^{\geq}) &\subseteq \underline{P}^{l'}(CI_t^{\geq}) \text{ and } \bar{P}^l(CI_t^{\geq}) \supseteq \bar{P}^{l'}(CI_t^{\geq}), \quad t=2, \dots, n, \\ \underline{P}^l(CI_t^{\leq}) &\subseteq \underline{P}^{l'}(CI_t^{\leq}) \text{ and } \bar{P}^l(CI_t^{\leq}) \supseteq \bar{P}^{l'}(CI_t^{\leq}), \quad t=1, \dots, n-1. \end{aligned}$$

The following two basic types of variable-consistency decision rules can be considered:

1. $D_{\geq}$ -decision rules with the following syntax:
 “if $\phi(x, q1) \geq r_{q1}$ and $\phi(x, q2) \geq r_{q2}$ and ...
 $\phi(x, qp) \geq r_{qp}$, then $x \in CI_t^{\geq}$ ” with confidence α (i.e., in fraction α of considered cases), where $P = \{q1, \dots, qp\} \subseteq C, (r_{q1}, \dots, r_{qp}) \in V_{q1} \times V_{q2} \times \dots \times V_{qp}$ and $t=2, \dots, n$;
2. $D_{\leq}$ -decision rules with the following syntax:
 “if $\phi(x, q1) \leq r_{q1}$ and $\phi(x, q2) \leq r_{q2}$ and ...
 $\phi(x, qp) \leq r_{qp}$, then $x \in CI_t^{\leq}$ ” with confidence α , where $P = \{q1, \dots, qp\} \subseteq C, (r_{q1}, \dots, r_{qp}) \in V_{q1} \times V_{q2} \times \dots \times V_{qp}$ and $t=1, \dots, n-1$.

The *variable-consistency* model is inspired by the *variable-precision* model proposed by Ziarko (1993, 1998) within the classical indiscernibility-based rough set approach.

Dominance-Based Rough Approximation of a Fuzzy Set

This section shows how the dominance-based rough set approach can be used for rough approximation of fuzzy sets.

A *fuzzy information base* is the 3-tuple $B = \langle U, F, \varphi \rangle$, where U is a finite set of *objects* (universe), $F = \{f_1, f_2, \dots, f_m\}$ is a finite set of *properties*, and $\varphi : U \times F \rightarrow [0, 1]$ is a function such that $\varphi(x, f_h) \in [0, 1]$ expresses the

credibility that object x has property f_h . Each object x from U is described by a vector

$$\text{Des}_F(x) = [\varphi(x, f_1), \dots, \varphi(x, f_m)],$$

called *description* of x in terms of the degrees to which it has properties from F ; it represents the available information about x . Obviously, $x \in U$ can be described in terms of any non-empty subset $E \subseteq F$ and in this case

$$\text{Des}_E(x) = [\varphi(x, f_h), f_h \in E].$$

For any $E \subseteq F$, the dominance relation D_E can be defined as follows: for all $x, y \in U$, x dominates y with respect to E (denotation $x D_E y$) if, for any $f_h \in E$,

$$\varphi(x, f_h) \geq \varphi(y, f_h).$$

Given $E \subseteq F$ and $x \in U$, let

$$\begin{aligned} D_E^+(x) &= \{y \in U : y D_E x\}, \\ D_E^-(x) &= \{y \in U : x D_E y\}. \end{aligned}$$

Let us consider a fuzzy set X in U , with its membership function $\mu_X : U \rightarrow [0, 1]$. For each cutting level $\alpha \in [0, 1]$ and for $* \in \{\geq, >\}$, the E -lower and the E -upper approximation of $X^{*\alpha} = \{y \in U : \mu_X(y) * \alpha\}$ with respect to $E \subseteq F$ (denotation $\underline{E}(X^{*\alpha})$ and $\bar{E}(X^{*\alpha})$, respectively), can be defined as:

$$\begin{aligned} \underline{E}(X^{*\alpha}) &= \{x \in U : D_E^+(x) \subseteq X^{*\alpha}\}, \\ \bar{E}(X^{*\alpha}) &= \{x \in U : D_E^-(x) \cap X^{*\alpha} \neq \emptyset\}. \end{aligned}$$

Rough approximations $\underline{E}(X^{*\alpha})$ and $\bar{E}(X^{*\alpha})$ can be expressed in terms of unions of granules $D_E^{\pm}(x)$ as follows:

$$\begin{aligned} \underline{E}(X^{*\alpha}) &= \cup_{x \in U} \{D_E^+(x) : D_E^+(x) \subseteq X^{*\alpha}\}, \\ \bar{E}(X^{*\alpha}) &= \cup_{x \in U} \{D_E^-(x) : D_E^-(x) \cap X^{*\alpha} \neq \emptyset\}. \end{aligned}$$

Analogously, for each cutting level $\alpha \in [0, 1]$ and for $\diamond \in \{\leq, <\}$, the E -lower and the E -upper

approximation of $X^{\diamond\alpha} = \{y \in U : \mu_X(y) \diamond \alpha\}$, with respect to $E \subseteq F$ (denotation $\underline{E}(X^{\diamond\alpha})$ and $\overline{E}(X^{\diamond\alpha})$, respectively), can be defined as:

$$\begin{aligned}\underline{E}(X^{\diamond\alpha}) &= \{x \in U : D_E^-(x) \subseteq X^{\diamond\alpha}\}, \\ \overline{E}(X^{\diamond\alpha}) &= \{x \in U : D_E^+(x) \cap X^{\diamond\alpha} \neq \emptyset\}.\end{aligned}$$

The rough approximations $\underline{E}(X^{\diamond\alpha})$ and $\overline{E}(X^{\diamond\alpha})$ can be expressed in terms of unions of granules $D_E^-(x)$ as follows:

$$\begin{aligned}\underline{E}(X^{\diamond\alpha}) &= \bigcup_{x \in U} \{D_E^-(x) : D_E^-(x) \subseteq X^{\diamond\alpha}\}, \\ \overline{E}(X^{\diamond\alpha}) &= \bigcup_{x \in U} \{D_E^-(x) : D_E^+(x) \cap X^{\diamond\alpha} \neq \emptyset\}.\end{aligned}$$

Let us remark that the rough approximations $\underline{E}(X^{\geq\alpha})$, $\overline{E}(X^{\geq\alpha})$, $\underline{E}(X^{\leq\alpha})$, and $\overline{E}(X^{\leq\alpha})$ can be rewritten as follows:

$$\begin{aligned}\underline{E}(X^{\geq\alpha}) &= \{x \in U : \forall w \in U, w D_E x \Rightarrow w \in X^{\geq\alpha}\}, \\ \overline{E}(X^{\geq\alpha}) &= \{x \in U : \exists w \in U \text{ such that } x D_E w \text{ and } w \in X^{\geq\alpha}\}, \\ \underline{E}(X^{\leq\alpha}) &= \{x \in U : \forall w \in U, x D_E w \Rightarrow w \in X^{\leq\alpha}\}, \\ \overline{E}(X^{\leq\alpha}) &= \{x \in U : \exists w \in U \text{ such that } w D_E x \text{ and } w \in X^{\leq\alpha}\}.\end{aligned}$$

Rough approximations $\underline{E}(X^{>\alpha})$, $\overline{E}(X^{>\alpha})$, $\underline{E}(X^{<\alpha})$, and $\overline{E}(X^{<\alpha})$ can be rewritten analogously by a simple replacement of “ $\geq$ ” with “ $>$ ” and “ $\leq$ ” with “ $<$.”

This reformulation of the rough approximations is concordant with the syntax of decision rules obtained in DRSA. For example, $\underline{E}(X^{\geq\alpha})$ is concordant with decision rules of the type

if object y has property f_{i1} to degree at least h_{i1} , and has property f_{i2} to degree at least h_{i2} , ..., and has property f_{ip} to degree at least h_{ip} , then object y belongs to set X to degree at least α ,

where $\{i1, \dots, ip\} = E$ and $h_{i1} = \varphi(x, f_{i1})$, ..., $h_{ip} = \varphi(x, f_{ip})$.

Let us remark that in the above approximations, even if $X^{\geq\alpha} = Y^{\leq\alpha}$, their approximations are, in general, different due to the different directions of cutting the membership functions of X and Y . Of course, a similar remark holds also for $X^{>\alpha}$ and $Y^{<\alpha}$. Considerations of the directions in the cuts $X^{\geq\alpha}$, $X^{>\alpha}$ and $X^{\leq\alpha}$, $X^{<\alpha}$ are important in the definition of the rough approximations of unions and intersections of cuts.

The rough approximations $\underline{E}(X^{\geq\alpha})$, $\overline{E}(X^{\geq\alpha})$, $\underline{E}(X^{\leq\alpha})$, $\overline{E}(X^{\leq\alpha})$ and $\underline{E}(X^{>\alpha})$, $\overline{E}(X^{>\alpha})$, $\underline{E}(X^{<\alpha})$, $\overline{E}(X^{<\alpha})$ satisfy the following inclusion properties: for any $0 \leq \alpha \leq 1$,

$$\begin{aligned}\underline{E}(X^{\geq\alpha}) &\subseteq X^{\geq\alpha} \subseteq \overline{E}(X^{\geq\alpha}), \\ \underline{E}(X^{\leq\alpha}) &\subseteq X^{\leq\alpha} \subseteq \overline{E}(X^{\leq\alpha}),\end{aligned}$$

$$\begin{aligned}\underline{E}(X^{>\alpha}) &\subseteq X^{>\alpha} \subseteq \overline{E}(X^{>\alpha}), \\ \underline{E}(X^{<\alpha}) &\subseteq X^{<\alpha} \subseteq \overline{E}(X^{<\alpha}).\end{aligned}$$

Furthermore, the following complementary properties hold: for any $0 \leq \alpha \leq 1$,

$$\begin{aligned}\underline{E}(X^{\geq\alpha}) &= U - \overline{E}(X^{<\alpha}), \\ \underline{E}(X^{\leq\alpha}) &= U - \overline{E}(X^{>\alpha}),\end{aligned}$$

$$\begin{aligned}\underline{E}(X^{>\alpha}) &= U - \overline{E}(X^{\leq\alpha}), \\ \underline{E}(X^{<\alpha}) &= U - \overline{E}(X^{\geq\alpha}).\end{aligned}$$

The following properties of monotonicity with respect to sets of properties also hold: for any $E_1 \subseteq E_2 \subseteq F$ and for any $0 \leq \alpha \leq 1$,

$$\underline{E}_1(X^{\geq\alpha}) \subseteq \underline{E}_2(X^{\geq\alpha}), \quad \underline{E}_1(X^{>\alpha}) \subseteq \underline{E}_2(X^{>\alpha}),$$

$$\underline{E}_1(X^{\leq\alpha}) \subseteq \underline{E}_2(X^{\leq\alpha}), \quad \underline{E}_1(X^{<\alpha}) \subseteq \underline{E}_2(X^{<\alpha}),$$

$$\begin{aligned}\bar{E}_1(X^{\geq \alpha}) &\supseteq \bar{E}_2(X^{\geq \alpha}), & \bar{E}_1(X^{> \alpha}) &\supseteq \bar{E}_2(X^{> \alpha}), \\ \bar{E}_1(X^{\leq \alpha}) &\supseteq \bar{E}_2(X^{\leq \alpha}), & \bar{E}_1(X^{< \alpha}) &\supseteq \bar{E}_2(X^{< \alpha}).\end{aligned}$$

One can consider also fuzzy rough approximations $\underline{X}_E^\uparrow$, $\underline{X}_E^\downarrow$, $\bar{X}_E^\uparrow$, $\bar{X}_E^\downarrow$, which are fuzzy sets with membership functions defined, respectively, as follows: for any $y \in U$,

$$\begin{aligned}\mu_{\underline{X}_E^\uparrow}(y) &= \max \{ \alpha \in [0, 1] : y \in \underline{E}(X^{\geq \alpha}) \}, \\ \mu_{\underline{X}_E^\downarrow}(y) &= \min \{ \alpha \in [0, 1] : y \in \underline{E}(X^{\leq \alpha}) \}, \\ \mu_{\bar{X}_E^\uparrow}(y) &= \max \{ \alpha \in [0, 1] : y \in \bar{E}(X^{\geq \alpha}) \}, \\ \mu_{\bar{X}_E^\downarrow}(y) &= \min \{ \alpha \in [0, 1] : y \in \bar{E}(X^{\leq \alpha}) \}.\end{aligned}$$

The membership function $\mu_{\underline{X}_E^\uparrow}(y)$ is defined as the upward lower fuzzy rough approximation of X with respect to E and can be interpreted in the following way. For any $\alpha, \beta \in [0, 1]$, $\alpha < \beta$ implies $X^{\geq \alpha} \supseteq X^{\geq \beta}$. Therefore, the greater the cutting level α , the smaller $X^{\geq \alpha}$ and, consequently, the smaller also its lower approximation $\underline{E}(X^{\geq \alpha})$. Thus, for each $y \in U$ and for each fuzzy set X , there is a threshold $k(y)$, $0 \leq k(y) \leq \mu_X(y)$, such that $y \in \underline{E}(X^{\geq \alpha})$ if $\alpha \leq k(y)$, and $y \notin \underline{E}(X^{\geq \alpha})$ if $\alpha > k(y)$. Since $k(y) = \mu_{\underline{X}_E^\uparrow}(y)$, this explains the interest of $\mu_{\underline{X}_E^\uparrow}(y)$. Analogous interpretation holds for $\mu_{\bar{X}_E^\uparrow}(y)$, defined as the membership function of the upward upper fuzzy rough approximation of X with respect to E .

The membership function $\mu_{\underline{X}_E^\downarrow}(y)$ is defined as the downward lower fuzzy rough approximation of X with respect to E and can be interpreted as follows. For any $\alpha, \beta \in [0, 1]$, $\alpha < \beta$ implies $X^{\leq \alpha} \subseteq X^{\leq \beta}$. Therefore, the greater the cutting level α , the greater $X^{\leq \alpha}$ and, consequently, its lower approximation $\underline{E}(X^{\leq \alpha})$. Thus, for each $y \in U$ and for each fuzzy set X , there is a threshold $h(y)$, $\mu_X(y) \leq h(y) \leq 1$, such that $y \in \underline{E}(X^{\leq \alpha})$ if $\alpha \geq h(y)$, and $y \notin \underline{E}(X^{\leq \alpha})$ if $\alpha < h(y)$. Observe that $h(y) = \mu_{\underline{X}_E^\downarrow}(y)$. Analogous interpretation holds for $\mu_{\bar{X}_E^\downarrow}(y)$, defined as the membership function of the downward upper fuzzy rough approximation of X with respect to E .

The membership functions of the upward and downward lower and upper fuzzy rough approximations can also be rewritten in the following

equivalent formulations, which has been proposed and investigated by Greco, Inuiguchi, and Słowiński (Greco et al. 2000b):

$$\begin{aligned}\mu_{\underline{X}_E^\uparrow}(y) &= \min \{ \mu_X(z) : z \in D_E^+(y) \}, & \mu_{\bar{X}_E^\uparrow}(y) &= \max \{ \mu_X(z) : z \in D_E^-(y) \}, \\ \mu_{\underline{X}_E^\downarrow}(y) &= \max \{ \mu_X(z) : z \in D_E^-(y) \}, & \mu_{\bar{X}_E^\downarrow}(y) &= \min \{ \mu_X(z) : z \in D_E^+(y) \}.\end{aligned}$$

The membership functions of the fuzzy rough approximations, $\mu_{\underline{X}_E^\uparrow}(y)$, $\mu_{\bar{X}_E^\uparrow}(y)$, $\mu_{\underline{X}_E^\downarrow}(y)$, and $\mu_{\bar{X}_E^\downarrow}(y)$, satisfy the following inclusion properties: for any $y \in U$,

$$\begin{aligned}\mu_{\underline{X}_E^\uparrow}(y) &\leq \mu_X(y) \leq \mu_{\bar{X}_E^\uparrow}(y), \\ \mu_{\underline{X}_E^\downarrow}(y) &\leq \mu_X(y) \leq \mu_{\bar{X}_E^\downarrow}(y).\end{aligned}$$

Furthermore, the following complementary properties hold: for any $y \in U$,

$$\mu_{\underline{X}_E^\uparrow}(y) = \mu_{\bar{X}_E^\downarrow}(y), \quad \mu_{\underline{X}_E^\downarrow}(y) = \mu_{\bar{X}_E^\uparrow}(y).$$

The following properties of monotonicity with respect to sets of properties also hold: for any $E_1 \subseteq E_2 \subseteq F$,

$$\begin{aligned}\mu_{\underline{X}_{E_1}^\uparrow}(y) &\leq \mu_{\underline{X}_{E_2}^\uparrow}(y), & \mu_{\underline{X}_{E_1}^\downarrow}(y) &\geq \mu_{\underline{X}_{E_2}^\downarrow}(y), \\ \mu_{\bar{X}_{E_1}^\uparrow}(y) &\geq \mu_{\bar{X}_{E_2}^\uparrow}(y), & \mu_{\bar{X}_{E_1}^\downarrow}(y) &\leq \mu_{\bar{X}_{E_2}^\downarrow}(y).\end{aligned}$$

Monotonic Rough Approximation of a Fuzzy Set Versus Classical Rough Set

What is the relationship between classical rough set and DRSA approximation of a fuzzy set? Greco et al. (2005, 2006b) proved that the former is a particular case of the latter, as shown below.

Let us remember that in classical rough set approach (Pawlak 1982, 1991), the original

information is expressed by means of an *information system*, that is the 4-tuple $\mathbf{S} = \langle U, Q, V, \phi \rangle$, where U is a finite set of *objects* (universe), $Q = \{q_1, q_2, \dots, q_m\}$ is a finite set of *attributes*, V_q is the set of values of the attribute q , $V = \bigcup_{q \in Q} V_q$ and $\phi : U \times Q \rightarrow V$ is a total function such that $\phi(x, q) \in V_q$ for each $q \in Q$, $x \in U$, called *information function*.

Therefore, each object x from U is described by a vector $Des_Q(x) = [\phi(x, q_1), \phi(x, q_2), \dots, \phi(x, q_m)]$, $(x, q_1), \phi(x, q_2), \dots, \phi(x, q_m)$, called *description* of x in terms of the evaluations of the attributes from Q ; it represents the available information about x . Obviously, $x \in U$ can be described in terms of any non-empty subset $P \subseteq Q$.

With every (non-empty) subset of attributes P there is associated an *indiscernibility relation* on U , denoted by I_P :

$$I_P = \{(x, y) \in U \times U : \phi(x, q) = \phi(y, q), \forall q \in P\}.$$

If $(x, y) \in I_P$ it is said that the objects x and y are P -indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric, and transitive). The family of all the equivalence classes of the relation I_P is denoted by U / I_P and the equivalence class containing an element $x \in U$ by $I_P(x)$, i.e.,

$$I_P(x) = \{y \in U : \phi(y, q) = \phi(x, q), \forall q \in P\}.$$

The equivalence classes of the relation I_P are called *P-elementary sets*.

Let $\mathbf{S}$ be an information system, X a non-empty subset of U and $\emptyset \neq P \subseteq Q$. The *P-lower approximation* and the *P-upper approximation* of X in $\mathbf{S}$ are defined, respectively, as:

$$\begin{aligned} \underline{P}(X) &= \{x \in U : I_P(x) \subseteq X\}, \\ \overline{P}(X) &= \{x \in U : I_P(x) \cap X \neq \emptyset\}. \end{aligned}$$

The elements of $\underline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation I_P contained in X ; the elements of $\overline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the

indiscernibility relation I_P containing at least one object x belonging to X . In other words, $\underline{P}(X)$ is the largest union of the P -elementary sets included in X , while $\overline{P}(X)$ is the smallest union of the P -elementary sets containing X .

Any information system can be expressed in terms of a specific type of an information base. An *information base* is called *Boolean* if $\varphi : U \times F \rightarrow \{0, 1\}$. A partition $F = \{F_1, \dots, F_r\}$ of the set of properties F , with $\text{card}(F_k) \geq 2$ for all $k = 1, \dots, r$, is called *canonical* if, for each $x \in U$ and for each $F_k \in F$, $k = 1, \dots, r$, there exists only one $f_j \in F_k$ for which $\varphi(x, f_j) = 1$ (and, therefore, for each $f_i \in F_k - \{f_j\}$, $\varphi(x, f_i) = 0$). The condition $\text{card}(F_k) \geq 2$ for all $k = 1, \dots, r$ is necessary because, otherwise, there would be at least one class $F_k = \{f'\}$ of the partition such that $\varphi(x, f') = 1$ for all $x \in U$, and this would mean that property f' gives no information and can be removed. Observe now that any *information system* $\mathbf{S} = \langle U, Q, V, \phi \rangle$ can be transformed to a Boolean information base $\mathbf{B} = \langle U, F, \varphi \rangle$ assigning to each $v \in V_q$, $q \in Q$, one property $f_{qv} \in F$ such that $\varphi(x, f_{qv}) = 1$ if $\phi(x, q) = v$, and $\varphi(x, f_{qv}) = 0$ otherwise. Let us remark that $F = \{F_1, \dots, F_r\}$, with $F_q = \{f_{qv}, v \in V_q\}$, $q \in Q$, is a canonical partition of F . The opposite transformation, from a Boolean information base to an information system, is not always possible, i.e., there may exist Boolean information bases which cannot be transformed into information systems, because their sets of properties do not admit any canonical partition, as shown by the following example.

Example Let us consider a Boolean information base $\mathbf{B}$, such that $U = \{x_1, x_2, x_3\}$, $F = \{f_1, f_2\}$, and function φ is defined by Table 1. One can see that $F = \{\{f_1, f_2\}\}$ is not a canonical partition because $\varphi(x_3, f_1) = \varphi(x_3, f_2) = 1$, while definition of canonical partition F does not allow that for an object $x \in U$, $\varphi(x, f_1) = \varphi(x, f_2) = 1$. Therefore, this Boolean information base has no equivalent information system. Let us remark that also the Boolean information base $\mathbf{B}'$ presented in Table 2, where $U = \{x_1, x_2, x_4\}$ and $F = \{f_1, f_2\}$, cannot be transformed to an information system,

Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach, Table 1 Information base $\mathbf{B}$

	f_1	f_2
x_1	0	1
x_2	1	0
x_3	1	1

Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach, Table 2 Information base $\mathbf{B}'$

	f_1	f_2
x_1	0	1
x_2	1	0
x_4	0	0

because partition $\mathbf{F} = \{\{f_1, f_2\}\}$ is not canonical. Indeed, $\varphi(x_4, f_1) = \varphi(x_4, f_2) = 0$, while definition of canonical partition $\mathbf{F}$ does not allow that for an object $x \in U$, $\varphi(x, f_1) = \varphi(x, f_2) = 0$.

The above says that consideration of rough approximation in the context of a Boolean information base is more general than the same consideration in the context of an information system. This means, of course, that the rough approximation considered in the context of a fuzzy information base is yet more general.

It is worth stressing that the Boolean information bases $\mathbf{B}$ and $\mathbf{B}'$ are not Boolean information systems. In fact, on one hand, a Boolean information base provides information about absence ($\varphi(x, f) = 0$) or presence ($\varphi(x, f) = 1$) of properties $f \in F$ in objects $x \in U$. On the other hand, a Boolean information system provides information about values assigned by attributes $q \in Q$, whose sets of values are $V_q = \{0, 1\}$, to objects $x \in U$, such that $\phi(x, q) = 1$ or $\phi(x, q) = 0$ for all $x \in U$ and $q \in Q$. Observe, therefore, that to transform a Boolean information system $\mathbf{S}$ into a Boolean information base $\mathbf{B}$, each attribute q of $\mathbf{S}$ corresponds to two properties f_{q0} and f_{q1} of $\mathbf{B}$, such that for all $x \in U$

- $\varphi(x, f_{q0}) = 1$ and $\varphi(x, f_{q1}) = 0$ if $\phi(x, q) = 0$,
- $\varphi(x, f_{q0}) = 0$ and $\varphi(x, f_{q1}) = 1$ if $\phi(x, q) = 1$.

Thus, the Boolean information base $\mathbf{B}$ in Table 1 and the Boolean information system $\mathbf{S}$ in Table 3 are different, despite they could seem identical. In fact, the Boolean information system $\mathbf{S}$ in Table 3 can be transformed into the Boolean information base $\mathbf{B}''$ in Table 4, which is clearly different from $\mathbf{B}$.



The equivalence between rough approximations in the context of a fuzzy information base and the classical definition of rough approximations in the context of an information system can be stated as follows (Greco et al. 2005, 2006b). Let us consider an information system and the corresponding Boolean information base; for each $P \subseteq Q$, let E^P be the set of all the properties corresponding to values v of attributes in P . Let X be a non-fuzzy (classical) set in U (i.e., $\mu_X: U \rightarrow \{0, 1\}$ and, therefore, for any $y \in U$, $\mu_X(y) = 1$ or $\mu_X(y) = 0$); then it is:

$$\underline{E}^P(X^{\geq 1}) = \underline{P}(X^{\geq 1}), \overline{E}^P(X^{\geq 1}) = \overline{P}(X^{\geq 1}),$$

$$\underline{E}^P(X^{\leq 0}) = \underline{P}(U - X^{\geq 1}),$$

$$\overline{E}^P(X^{\leq 0}) = \overline{P}(U - X^{\geq 1}).$$

This result proves that the rough approximation of a non-fuzzy set X in a Boolean information base admitting a canonical partition is equivalent to the classical rough approximation of set X in the corresponding information system. Therefore, the classical rough approximation is a particular case of the dominance-based rough approximation in a fuzzy information base.

Dominance-Based Rough Set Approach to Case-Based Reasoning

This section presents rough approximation of a fuzzy set using a similarity relation in the context of case-based reasoning (Greco et al. 2006a; Szelag et al. 2011, 2016).

Case-based reasoning (for a general introduction to case-based reasoning see, e.g., Kolodner

Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach, Table 3 Information system $\mathcal{S}$

	q_1	q_2
x_1	0	1
x_2	1	0
x_3	1	1

Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach, Table 4 Information base $\mathcal{B}'$

	$f_{q_1 0}$	$f_{q_1 1}$	$f_{q_2 0}$	$f_{q_2 1}$
x_1	1	0	0	1
x_2	0	1	1	0
x_3	0	1	0	1

(1993); for a fuzzy set approach to case-based reasoning see Dubois et al. (1998)) is a paradigm in machine learning whose idea is that a new problem can be solved by noticing its similarity to a set of problems previously solved. Case-based reasoning regards the inference of some proper conclusions related to a new situation by the analysis of similar cases from a memory of previous cases. It is based on two principles (Leake 1996):

- (a) Similar problems have similar solutions.
- (b) Types of encountered problems tend to recur.

Gilboa and Schmeidler (2001) observed that the basic idea of case-based reasoning can be found in the following sentence of Hume (1748): “From causes which appear *similar* we expect similar effects. This is the sum of all our experimental conclusions.” Rephrasing Hume, one can say that “the more similar are the causes, the more similar one expects the effects.” Therefore, measuring similarity is the essential point of all case-based reasoning and, particularly, of fuzzy set approach to case-based reasoning (Dubois et al. 1998). This explains the many problems that measuring similarity generates within case-based reasoning. Problems of modeling similarity are relative to two levels:

- At the level of similarity with respect to single features: how to define a meaningful similarity measure with respect to a single feature?
- At the level of similarity with respect to all features: how to properly aggregate the similarity measures with respect to single features in order to obtain a comprehensive similarity measure?

For the above reasons, Greco et al. (2006a) proposes a DRSA approach to case-based reasoning, which tries to be possibly “neutral” and “objective” with respect to similarity relation. At the level of similarity concerning single features, the DRSA approach to case-based reasoning considers only ordinal properties of similarity, and at the level of aggregation, it does not impose any particular functional aggregation based on some very specific axioms (see, e.g., Gilboa and Schmeidler (2001)), but it considers a set of decision rules based on the general monotonicity property of comprehensive similarity with respect to similarity of single features. Therefore, the DRSA approach to case-based reasoning is very little “invasive,” comparing to the many other existing approaches.

Let us consider a *pairwise fuzzy information base* being the 3-tuple

$$\mathcal{B} = \langle U, F, \sigma \rangle,$$

where U is a finite set of *objects* (universe), $F = \{f_1, f_2, \dots, f_m\}$ is a finite set of *features*, and $\sigma : U \times U \times F \rightarrow [0, 1]$ is a function such that $\sigma(x, y, f_h) \in [0, 1]$ expresses the credibility that object x is similar to object y w.r.t. feature f_h . The minimal requirement function σ must satisfy that, for all $x \in U$ and for all $f_h \in F$, $\sigma(x, x, f_h) = 1$. Therefore, each pair of objects $(x, y) \in U \times U$ is described by a vector

$$\text{Des}_F(x, y) = [\sigma(x, y, f_1), \dots, \sigma(x, y, f_m)],$$

called *description* of (x, y) in terms of the credibilities of similarity with respect to features from F ; it represents the available information about similarity between x and y . Obviously, similarity

between x and y , $x, y \in U$, can be described in terms of any non-empty subset $E \subseteq F$ as follows:

$$\text{Des}_E(x, y) = [\sigma(x, y, f_h), f_h \in E].$$

With respect to any $E \subseteq F$, the dominance relation D_E can be defined on $U \times U$ as follows: for any $x, y, w, z \in U$, (x, y) dominates (w, z) with respect to E (denotation $(x, y)D_E(w, z)$) if, for any $f_h \in E$,

$$\sigma(x, y, f_h) \geq \sigma(w, z, f_h).$$

Given $E \subseteq F$ and $x, y \in U$, let

$$D_E^+(y, x) = \{w \in U : (w, x)D_E(y, x)\},$$

$$D_E^-(y, x) = \{w \in U : (y, x)D_E(w, x)\}.$$

In the pair (y, x) , x is considered to be a *reference object*, while y can be called a *limit object*, because it is conditioning the membership of w in $D_E^+(y, x)$ and in $D_E^-(y, x)$.

For each $x \in U$ and $\alpha \in [0, 1]$ and $* \in \{\geq, >\}$, the lower approximation of $X^{*\alpha}$, $\underline{E}_\sigma(X^{*\alpha})$, and the upper approximation of $X^{*\alpha}$, $\bar{E}_\sigma(X^{*\alpha})$, based on similarity σ with respect to $E \subseteq F$ and x , respectively, can be defined as:

$$\underline{E}(x)_\sigma(X^{*\alpha}) = \{y \in U : D_E^+(y, x) \subseteq X^{*\alpha}\},$$

$$\bar{E}(x)_\sigma(X^{*\alpha}) = \{y \in U : D_E^-(y, x) \cap X^{*\alpha} \neq \emptyset\}.$$

For the sake of simplicity, in the following, only $\underline{E}(x)_\sigma(X^{\geq\alpha})$ and $\bar{E}(x)_\sigma(X^{\geq\alpha})$ with $x \in X^{\geq\alpha}$ are considered. Of course, analogous considerations hold for $\underline{E}(x)_\sigma(X^{>\alpha})$ and $\bar{E}(x)_\sigma(X^{>\alpha})$. Observe that the lower approximation of $X^{\geq\alpha}$ with respect to x contains all the objects $y \in U$ such that any object w , being similar to x at least as much as y is similar to x w.r.t. all the considered features $E \subseteq F$, also belongs to $X^{\geq\alpha}$. Thus, the data from the fuzzy pairwise information base $\mathbf{B}$ confirm that if w is similar to x not less than $y \in \underline{E}(x)_\sigma(X^{\geq\alpha})$ is similar to x w.r.t. all the considered features $E \subseteq F$, then w belongs to $X^{\geq\alpha}$. In other words, x is a reference object and $y \in \underline{E}(x)_\sigma(X^{\geq\alpha})$ is a limit object which

belongs “certainly” to set X with credibility at least α ; the limit is understood such that all objects w that are similar to x w.r.t. considered features at least as much as y is similar to x , also belong to X with credibility at least α .

Analogously, the upper approximation of $X^{\geq\alpha}$ with respect to x contains all objects $y \in U$ such that there is at least one object w , being similar to x at most as much as y is similar to x w.r.t. all the considered features $E \subseteq F$, which belongs to $X^{\geq\alpha}$. Thus, the data from the fuzzy pairwise information base $\mathbf{B}$ confirm that if w is similar to x not less than $y \in \bar{E}(x)_\sigma(X^{\geq\alpha})$ is similar to x w.r.t. all the considered features $E \subseteq F$, then it is possible that w belongs to $X^{\geq\alpha}$. In other words, x is a reference object and $y \in \bar{E}(x)_\sigma(X^{\geq\alpha})$ is a limit object which belongs “possibly” to set X with credibility at least α ; the limit is understood such that all objects $z \in U$ similar to x not less than y w.r.t. considered features, possibly belong to $X^{\geq\alpha}$.

For each $x \in U$ and $\alpha \in [0, 1]$ and $\diamond \in \{\leq, <\}$, the lower approximation of $X^{\diamond\alpha}$, $\underline{E}(x)_\sigma(X^{\diamond\alpha})$, and the upper approximation of $X^{\diamond\alpha}$, $\bar{E}(x)_\sigma(X^{\diamond\alpha})$, based on similarity σ with respect to $E \subseteq F$ and x , respectively, can be defined as:

$$\underline{E}(x)_\sigma(X^{\diamond\alpha}) = \{y \in U : D_E^-(y, x) \subseteq X^{\diamond\alpha}\},$$

$$\bar{E}(x)_\sigma(X^{\diamond\alpha}) = \{y \in U : D_E^+(y, x) \cap X^{\diamond\alpha} \neq \emptyset\}.$$

For the sake of simplicity, in the following, only $\underline{E}(x)_\sigma(X^{\leq\alpha})$ and $\bar{E}(x)_\sigma(X^{\leq\alpha})$ with $x \in X^{\leq\alpha}$ are considered. Of course, analogous considerations hold for $\underline{E}(x)_\sigma(X^{<\alpha})$ and $\bar{E}(x)_\sigma(X^{<\alpha})$. Observe that the lower approximation of $X^{\leq\alpha}$ with respect to x contains all the objects $y \in U$ such that any object w , being similar to x at most as much as y is similar to x w.r.t. all the considered features $E \subseteq F$, also belongs to $X^{\leq\alpha}$. Thus, the data from the fuzzy pairwise information base $\mathbf{B}$ confirm that if w is similar to x not more than $y \in \underline{E}(x)_\sigma(X^{\leq\alpha})$ is similar to x w.r.t. all the considered features $E \subseteq F$, then w belongs to $X^{\leq\alpha}$. In other words, x is a reference object and $y \in \underline{E}(x)_\sigma(X^{\leq\alpha})$ is a limit object which belongs “certainly” to set X with credibility at most α ; the limit is understood such that all objects w that are similar to x w.r.t. considered features at most as

much as y is similar to x , also belong to X with credibility at most α .

Analogously, the upper approximation of $X^{\leq\alpha}$ with respect to x contains all the objects $y \in U$ such that there is at least one object w , being similar to x at least as much as y is similar to x with respect to all the considered features $E \subseteq F$, which belongs to $X^{\leq\alpha}$. Thus, the data from the fuzzy pairwise information base B confirm that if w is similar to x not more than $y \in \bar{E}(x)_\sigma(X^{\leq\alpha})$ is similar to x w.r.t. all the considered features $E \subseteq F$, then it is possible that w belongs to $X^{\leq\alpha}$. In other words, x is a reference object and $y \in \bar{E}_\sigma(X^{\leq\alpha})$ is a limit object which belongs “possibly” to set X with credibility at most α ; the limit is understood such that all objects $z \in U$ similar to x not more than y w.r.t. considered features, possibly belong to $X^{\leq\alpha}$.

Observe that the rough approximations $\underline{E}(x)_\sigma(X^{\geq\alpha})$, $\bar{E}(x)_\sigma(X^{\geq\alpha})$, $\underline{E}(x)_\sigma(X^{\leq\alpha})$, and $\bar{E}(x)_\sigma(X^{\leq\alpha})$ can be rewritten as

$$\begin{aligned} \underline{E}(x)_\sigma(X^{\geq\alpha}) \\ = \{y \in U : \forall w \in U, (w, x)D_E(y, x) \Rightarrow w \in X^{\geq\alpha}\}, \end{aligned}$$

$$\begin{aligned} \bar{E}(x)_\sigma(X^{\geq\alpha}) \\ = \{y \in U : \exists w \in U \text{ such that } (y, x)D_E(w, x) \text{ and } w \in X^{\geq\alpha}\}, \end{aligned}$$

$$\begin{aligned} \underline{E}(x)_\sigma(X^{\leq\alpha}) \\ = \{y \in U : \forall w \in U, (y, x)D_E(w, x) \Rightarrow w \in X^{\leq\alpha}\}, \end{aligned}$$

$$\begin{aligned} \bar{E}(x)_\sigma(X^{\leq\alpha}) \\ = \{y \in U : \exists w \in U \text{ such that } (w, x)D_E(y, x) \text{ and } w \in X^{\leq\alpha}\}. \end{aligned}$$

This formulation of the rough approximation is concordant with the syntax of the decision rules induced by means of DRSA from a fuzzy pairwise information base. More precisely,

- $\underline{E}(x)_\sigma(X^{\geq\alpha})$ is concordant with decision rules of the type:
“if object w is similar to object x w.r.t. feature f_{i1} to degree at least h_{i1} and w.r.t. feature f_{i2} to degree at least h_{i2} and ... and w.r.t. feature f_{ip} to degree at least h_{ip} , then object w belongs to set X to degree at least α ,”

- $\bar{E}(x)_\sigma(X^{\geq\alpha})$ is concordant with decision rules of the type:

“if object w is similar to object x w.r.t. feature f_{i1} to degree at least h_{i1} and w.r.t. feature f_{i2} to degree at least h_{i2} and ... and w.r.t. feature f_{ip} to degree at least h_{ip} , then object w could belong to set X to degree at least α ,”

- $\underline{E}(x)_\sigma(X^{\leq\alpha})$ is concordant with decision rules of the type:

“if object w is similar to object x w.r.t. feature f_{i1} to degree at most h_{i1} and w.r.t. feature f_{i2} to degree at most h_{i2} and ... and w.r.t. feature f_{ip} to degree at most h_{ip} , then object w belongs to set X to degree at most α ,”

- $\bar{E}(x)_\sigma(X^{\leq\alpha})$ is concordant with decision rules of the type:

“if object w is similar to object x w.r.t. feature f_{i1} to degree at most h_{i1} and w.r.t. feature f_{i2} to degree at most h_{i2} and ... and w.r.t. feature f_{ip} to degree at most h_{ip} , then object w could belong to set X to degree at most α ,”

where $\{i1, \dots, ip\} = E$ and $h_{i1}, \dots, h_{ip} \in [0, 1]$.

The above definitions of rough approximations and the syntax of decision rules are based on ordinal properties of similarity relations only. In fact, no algebraic operation, such as sum or product, involving cardinal properties of function σ measuring credibility of similarity relations is considered. This is an important characteristic of our approach in comparison with alternative approaches to case-based reasoning.

Let us remark that, similarly to DRSA approximation in an information base, in the case of DRSA approximation in a fuzzy pairwise information base, even if for two fuzzy sets X and Y it is $X^{\geq\alpha} = Y^{\leq\alpha}$, their approximations may be different due to the different directions of cutting the membership functions of sets X and Y .

Rough approximations in a fuzzy pairwise information base satisfy the following interesting properties.

Theorem 1 Given a fuzzy pairwise information base $B = \langle U, F, \sigma \rangle$ and a fuzzy set X in U with membership function $\mu_X(\cdot)$, the following properties hold for any $E \subseteq F$:

1. For any $0 \leq \alpha \leq 1$,

$$\underline{E}(x)_\sigma(X^{\leq \alpha}) \subseteq X^{\leq \alpha} \subseteq \overline{E}(x)_\sigma(X^{\leq \alpha}),$$

$$\underline{E}(x)_\sigma(X^{\geq \alpha}) \subseteq X^{\geq \alpha} \subseteq \overline{E}(x)_\sigma(X^{\geq \alpha}),$$

$$\underline{E}(x)_\sigma(X^{< \alpha}) \subseteq X^{< \alpha} \subseteq \overline{E}(x)_\sigma(X^{< \alpha}),$$

$$\underline{E}(x)_\sigma(X^{> \alpha}) \subseteq X^{> \alpha} \subseteq \overline{E}(x)_\sigma(X^{> \alpha}).$$

2. For any $0 \leq \alpha \leq 1$,

$$\begin{aligned} \underline{E}(x)_\sigma(X^{\leq \alpha}) &= U - \overline{E}(x)_\sigma(X^{> \alpha}), \underline{E}(x)_\sigma(X^{\geq \alpha}) \\ &= U - \overline{E}(x)_\sigma(X^{< \alpha}). \end{aligned}$$

3. For any $0 \leq \alpha \leq \beta \leq 1$,

$$\begin{aligned} \underline{E}(x)_\sigma(X^{\leq \alpha}) &\subseteq \underline{E}(x)_\sigma(X^{\leq \beta}), \quad \underline{E}(x)_\sigma(X^{< \alpha}) \\ &\subseteq \underline{E}(x)_\sigma(X^{< \beta}), \end{aligned}$$

$$\begin{aligned} \underline{E}(x)_\sigma(X^{\geq \alpha}) &\supseteq \underline{E}(x)_\sigma(X^{\geq \beta}), \quad \underline{E}(x)_\sigma(X^{> \alpha}) \\ &\supseteq \underline{E}(x)_\sigma(X^{> \beta}), \end{aligned}$$

$$\begin{aligned} \overline{E}(x)_\sigma(X^{\leq \alpha}) &\subseteq \overline{E}(x)_\sigma(X^{\leq \beta}), \quad \overline{E}(x)_\sigma(X^{< \alpha}) \\ &\subseteq \overline{E}(x)_\sigma(X^{< \beta}), \end{aligned}$$

$$\begin{aligned} \overline{E}(x)_\sigma(X^{\geq \alpha}) &\supseteq \overline{E}(x)_\sigma(X^{\geq \beta}), \quad \overline{E}(x)_\sigma(X^{> \alpha}) \\ &\supseteq \overline{E}(x)_\sigma(X^{> \beta}). \end{aligned}$$

4. For any $x, y, w, z \in U$ and for any $0 \leq \alpha \leq 1$,

$$\begin{aligned} [(y, x)D_E(w, x) \text{ and } w \in \underline{E}(x)_\sigma(X^{\geq \alpha})] \\ \Rightarrow y \in \underline{E}(x)_\sigma(X^{\geq \alpha}), \end{aligned}$$

$$\begin{aligned} [(y, x)D_E(w, x) \text{ and } w \in \underline{E}(x)_\sigma(X^{> \alpha})] \\ \Rightarrow y \in \underline{E}(x)_\sigma(X^{> \alpha}), \end{aligned}$$

$$\begin{aligned} [(y, x)D_E(w, x) \text{ and } w \in \overline{E}(x)_\sigma(X^{\geq \alpha})] \\ \Rightarrow y \in \overline{E}(x)_\sigma(X^{\geq \alpha}), \end{aligned}$$

$$\begin{aligned} [(y, x)D_E(w, x) \text{ and } w \in \overline{E}(x)_\sigma(X^{> \alpha})] \\ \Rightarrow y \in \overline{E}(x)_\sigma(X^{> \alpha}), \end{aligned}$$

$$\begin{aligned} [(w, x)D_E(y, x) \text{ and } w \in \underline{E}(x)_\sigma(X^{\leq \alpha})] \\ \Rightarrow y \in \underline{E}(x)_\sigma(X^{\leq \alpha}), \end{aligned}$$

$$\begin{aligned} [(w, x)D_E(y, x) \text{ and } w \in \underline{E}(x)_\sigma(X^{< \alpha})] \\ \Rightarrow y \in \underline{E}(x)_\sigma(X^{< \alpha}), \end{aligned}$$

$$\begin{aligned} [(w, x)D_E(y, x) \text{ and } w \in \overline{E}(x)_\sigma(X^{\leq \alpha})] \\ \Rightarrow y \in \overline{E}(x)_\sigma(X^{\leq \alpha}), \end{aligned}$$

$$\begin{aligned} [(w, x)D_E(y, x) \text{ and } w \in \overline{E}(x)_\sigma(X^{< \alpha})] \\ \Rightarrow y \in \overline{E}(x)_\sigma(X^{< \alpha}). \end{aligned}$$

5. For any $E_1 \subseteq E_2 \subseteq F$ and for any $0 \leq \alpha \leq 1$,

$$\begin{aligned} \underline{E}_1(x)_\sigma(X^{\leq \alpha}) &\subseteq \underline{E}_2(x)_\sigma(X^{\leq \alpha}), \quad \underline{E}_1(x)_\sigma(X^{< \alpha}) \\ &\subseteq \underline{E}_2(x)_\sigma(X^{< \alpha}), \end{aligned}$$

$$\begin{aligned} \underline{E}_1(x)_\sigma(X^{\geq \alpha}) &\subseteq \underline{E}_2(x)_\sigma(X^{\geq \alpha}), \quad \underline{E}_1(x)_\sigma(X^{> \alpha}) \\ &\subseteq \underline{E}_2(x)_\sigma(X^{> \alpha}), \end{aligned}$$

$$\begin{aligned} \overline{E}_1(x)_\sigma(X^{\leq \alpha}) &\supseteq \overline{E}_2(x)_\sigma(X^{\leq \alpha}), \quad \overline{E}_1(x)_\sigma(X^{< \alpha}) \\ &\supseteq \overline{E}_2(x)_\sigma(X^{< \alpha}), \end{aligned}$$

$$\begin{aligned} \overline{E}_1(x)_\sigma(X^{\geq \alpha}) &\supseteq \overline{E}_2(x)_\sigma(X^{\geq \alpha}), \quad \overline{E}_1(x)_\sigma(X^{> \alpha}) \\ &\supseteq \overline{E}_2(x)_\sigma(X^{> \alpha}). \end{aligned}$$

An Algebraic Structure for Dominance-Based Rough Set Approach

A plurality of formal representations of classical rough set theory has been proposed in terms of abstract algebraic structures (see, e.g., Chap. 12 in Polkowski (2002) or Cattaneo and Ciucci (2004)). This section presents an algebraic characterization of DRSA in terms of bipolar Pawlak-Brouwer-Zadeh lattice (Greco et al. 2012b). The bipolar Pawlak-Brouwer-Zadeh lattice is based on the

Brouwer-Zadeh lattice (Cattaneo and Nisticò 1989) that was introduced as an algebraic structure permitting representation of possibility and necessity in presence of vagueness. It proved to be an interesting abstract model for classical rough set theory (Cattaneo 1997; Cattaneo and Ciucci 2004). In the Brouwer-Zadeh lattice, the elements of the lattice represent the pairs (I, E) where I and E are the lower approximation (interior) and the complement of the upper approximation (exterior) of a given set X , respectively.

The Brouwer-Zadeh lattice has been extended to represent DRSA with the bipolar complemented de Morgan Brouwer-Zadeh lattice (Greco et al. 2012a). In this case, the elements of the lattice can be “positive” concepts or “negative” concepts. For example, in a problem of evaluation of students, the concept of “good students” is “positive,” while the concept of “bad students” is “negative.” Consequently, the positive concept “good students” is represented by the pair (I_G, E_G) , where I_G is the dominance-based rough approximation of students “at least good,” that is, the set of students “certainly good” and E_G is the dominance-based rough approximation of students “not good or worse,” that is, the set of students “certainly not good.” Analogously, the negative concept “bad students” is represented by the pair (I_B, E_B) , where I_B is the dominance-based rough approximation of students “at most bad,” that is, the set of students “certainly bad” and E_B is the dominance-based rough approximation of students “not bad or better,” that is, the set of students “certainly not bad.”

In Greco et al. (2007), we proposed another extension of the Brouwer-Zadeh lattice for indiscernibility-based rough set approach, called Pawlak-Brouwer-Zadeh lattice, where the basic vagueness is represented by a pair (A, B) , with A being the necessity kernel and B being the non-possibility kernel, and ambiguity due to the coarseness of information is introduced through a new operator, called Pawlak operator. Pawlak operator assigns a pair (C, D) to the pair (A, B) , with $A \cap B = \emptyset$, such that C and D are the lower approximations of A and B , respectively.

The bipolar complemented de Morgan Brouwer-Zadeh lattice puts together the de

Morgan Brouwer-Zadeh lattice and the Pawlak-Brouwer-Zadeh lattice in order to get an algebraic model permitting to distinguish vagueness from ambiguity in DRSA. The bipolar Pawlak-Brouwer-Zadeh lattice is mainly based on the bipolar Pawlak operator that, to a pair (A, B) representing a positive concept, assigns the pair (C, D) , where C is the DRSA lower approximation of A being an upward union of classes and D is the DRSA lower approximation of B being a downward union of classes. For example, if (A, B) is the concept “good students,” so that A is the set of students evaluated as good and B is the set of students evaluated as bad, then C is the DRSA lower approximation of students “at least good” and D is the DRSA lower approximation of students “at most bad.”

Formally, the bipolar Pawlak-Brouwer-Zadeh lattice can be described as follows. A system $\langle \Sigma, \wedge, \vee, ', \sim, 0, 1 \rangle$ is a quasi-Brouwer-Zadeh distributive lattice (Cattaneo and Nisticò 1989) if the following properties (1)–(4) hold:

- (1) Σ is a distributive lattice with respect to the join and the meet operations $\vee, \wedge$ whose induced partial order relation is

$$a \leq b \text{ iff } a = a \wedge b \text{ (equivalently } b = a \vee b).$$

Moreover, it is required that Σ is bounded by the least element 0 and the greatest element 1:

$$\forall a \in \Sigma, \quad 0 \leq a \leq 1.$$

- (2) The unary operation $' : \Sigma \rightarrow \Sigma$ is a Kleene (also Zadeh or fuzzy) complementation. In other words, for arbitrary $a, b \in \Sigma$,

$$(K1) \quad a'' = a,$$

$$(K2) \quad (a \vee b)' = a' \wedge b',$$

$$(K3) \quad a \wedge a' \leq b \vee b'.$$

- (3) The unary operation $\sim : \Sigma \rightarrow \Sigma$ is a Brouwer (or intuitionistic) complementation. In other words, for arbitrary $a, b \in \Sigma$,

$$(B1) \quad a \wedge a^{\sim\sim} = a,$$

$$(B2) (a \vee b)^\sim = a^\sim \wedge b^\sim,$$

$$(B3) a \wedge a^\sim = 0.$$

- (4) The two complementations are linked by the interconnection rule which must hold for arbitrary $a \in \Sigma$:

$$(\text{in}) a^\sim \leq a'.$$

A structure $\langle \Sigma, \wedge, \vee, ', \sim, 0, 1 \rangle$ is a Brouwer-Zadeh distributive lattice if it is a quasi-Brouwer-Zadeh distributive lattice satisfying the stronger interconnection rule:

$$(\text{s-in}) a^{\sim\sim} = a'^{\sim}.$$

A Brouwer-Zadeh distributive lattice satisfying also the $\vee$ de Morgan property

$$(B2a)(a \wedge b)^\sim = a^\sim \vee b^\sim$$

is called a de Morgan Brouwer-Zadeh distributive lattice.

A bipolar de Morgan Brouwer-Zadeh distributive lattice is a system $\langle \Sigma, \Sigma^+, \wedge, \vee, ', \sim, 0, 1 \rangle$ for which the following properties hold

(B1) $\langle \Sigma, \wedge, \vee, ', \sim, 0, 1 \rangle$ is a de Morgan Brouwer-Zadeh distributive lattice,

(B2) $\Sigma^+ \subseteq \Sigma$ is a distributive lattice with respect to the join and the meet operations $\vee$ and $\wedge$, and

(B3) $a \in \Sigma^+$ implies that $a'^\sim, a'^{\sim} \in \Sigma^+$.

Observe that by (B3), (K1), and (s-in), we have that $a \in \Sigma^+$ implies $a'', a^{\sim\sim} \in \Sigma^+$. Let us consider the set

$$\Sigma^- = \{a \in \Sigma : a = b' \text{ or } a = b^\sim \text{ with } b \in \Sigma\}.$$

Σ^- is a distributive lattice with respect to the join and the meet operations $\vee$ and $\wedge$, and $a \in \Sigma^-$ implies that $a'^\sim, a'^{\sim}, a^{\sim\sim}, a'' \in \Sigma^-$.

A bipolar approximation operator, called bipolar Pawlak operator, on a bipolar de Morgan Brouwer-Zadeh distributive lattice is a unary operation $^M : \Sigma^+ \rightarrow \Sigma^+$ satisfying the following properties:

- $a \leq b$ implies $b^{M\sim} \leq a^{M\sim}$.
- $a \leq b$ implies $a^{M\sim} \leq b^{M\sim}$.
- $a^{M\sim} \leq a^\sim$.
- $a^{M\sim} \leq a'^\sim$.
- $0^M = 0$.
- $1^M = 1$.
- $a^\sim = b^\sim$ implies $a^M \wedge b^M = (a \wedge b)^M$.
- $a'^\sim = b'^\sim$ implies $a^M \vee b^M = (a \vee b)^M$.
- $a^M \vee b^M \leq (a \vee b)^M$.
- $(a \wedge b)^M \leq a^M \wedge b^M$.
- $a^{MM} = a^M$.
- $a^{M\sim M} = a^{M\sim}$.
- $a^{M\sim M} = a'^{\sim M}$.
- $(a^M \wedge b^M)^M = a^M \wedge b^M$.
- $(a^M \vee b^M)^M = a^M \vee b^M$.

An ordered knowledge base $K = (U, R)$ is a relational system where the universe $U \neq \emptyset$ is a finite set and R is a preorder on U , that is, R is a reflexive and transitive binary relation on U . For any $x \in U$, let us consider its dominating set $[x]_R^\geq = \{y \in U : yRx\}$ and its dominated set $[x]_R^\leq = \{y \in U : xRy\}$. Given the ordered knowledge base $K = (U, R)$, to each subset $X \subseteq U$, one can associate the four subsets $\underline{R}^\geq X$, $\overline{R}^\geq X$, $\underline{R}^\leq X$, and $\overline{R}^\leq X$:

$$\underline{R}^\geq X = \{x \in U : [x]_R^\geq \subseteq X\},$$

$$\overline{R}^\geq X = \{x \in U : [x]_R^\leq \cap X \neq \emptyset\},$$

$$\underline{R}^\leq X = \{x \in U : [x]_R^\leq \subseteq X\},$$

$$\overline{R}^\leq X = \{x \in U : [x]_R^\geq \cap X \neq \emptyset\}.$$

$\underline{R}^\geq X$ and $\overline{R}^\geq X$ are called the upward lower and upper approximations of X , while $\underline{R}^\leq X$ and $\overline{R}^\leq X$ are called the downward lower and upper approximations of X . Of course, the preorder R can be identified with the dominance relation so that $\underline{R}^\geq X$, $\overline{R}^\geq X$, $\underline{R}^\leq X$, and $\overline{R}^\leq X$ can be straightforwardly identified with dominance-based rough approximations.

Given an ordered knowledge base $K = (U, R)$, let us consider $\mathcal{U}^+ \subseteq 2^U$, such that

- $\emptyset, U \in \mathcal{U}^+$.
- For all $A, B \subseteq U$, if $A, B \in \mathcal{U}^+$ then also $A \cap B, A \cup B \in \mathcal{U}^+$.
- For all $A \subseteq U$, if $A \in \mathcal{U}^+$ then also $\underline{R}^{\geq} A, \overline{R}^{\geq} A \in \mathcal{U}^+$.

The sets $A \subseteq U$, such that $A \in \mathcal{U}^+$, are called positive sets. Let us observe that, possibly, $\mathcal{U}^+ = 2^U$, in which case all subsets of U are positive. Having a family $\mathcal{U}^+$ of positive sets, one can define a family $\mathcal{U}^-$ of negative sets as follows:

$$\mathcal{U}^- = \{A \subseteq U : U - A \in \mathcal{U}^+\}.$$

On the basis of $\mathcal{U}^+$ (and, consequently, of $\mathcal{U}^-$), one can further define the set $\mathcal{W}(\mathcal{U}^+)$ of all pairs $\langle A, B \rangle$, such that $A \in \mathcal{U}^+, B \in \mathcal{U}^-$, and $A \cap B = \emptyset$. Let us observe that if $\mathcal{U}^+ = 2^U$, then $\mathcal{W}(\mathcal{U}^+) = 3^U$.

Given an ordered knowledge base $K = (U, R)$ and a family of positive sets $\mathcal{U}^+$ (and, therefore, $\mathcal{W}(\mathcal{U}^+)$), we can define an unary operator $N : \mathcal{W}(\mathcal{U}^+) \rightarrow \mathcal{W}(\mathcal{U}^+)$, as follows: for any $\langle A, B \rangle \in \mathcal{W}(\mathcal{U}^+)$

$$\langle A, B \rangle^N = \langle \underline{R}^{\geq} A, \underline{R}^{\leq} B \rangle.$$

In Greco et al. (2012b), we proved that the structure $\langle 3^U, \mathcal{W}(\mathcal{U}^+), \sqcap, \sqcup, -, \approx, N, \langle \emptyset, U \rangle, \langle U, \emptyset \rangle \rangle$ is a bipolar Pawlak-Brouwer-Zadeh lattice, and thus it is an abstract algebraic model for DRSA.

Conclusions

This article provides arguments for the claim that the dominance-based rough set approach (DRSA) is a proper way of handling monotonically ordered data in granular computing. Referring to some ideas of Leibniz, Frege, Boole, and Łukasiewicz, DRSA represents fundamental concepts of rough set theory in terms of a generalization that takes into account ordinal properties of data, permitting to deal with the graduality of fuzzy sets. DRSA in the context of ordinal classification, its fuzzy extension, and a rough probabilistic model of DRSA (variable-consistency dominance-based rough set approach,

VC-DRSA) has been presented. Moreover, dominance-based rough approximation of a fuzzy set that infers the most cautious conclusion from available imprecise information has been discussed; otherwise than almost all known fuzzy rough set approaches, the dominance-based rough approximation of a fuzzy set does not require any fuzzy connective which is always arbitrary to some extent.

Knowledge induced from dominance-based rough approximations of fuzzy sets, or more generally, from ordinal data, is represented in terms of gradual decision rules. The dominance-based rough approximations of fuzzy sets generalize the classical rough approximations of crisp sets, as proved by showing that the classical rough set approach is one of its particular cases. Due to considering only ordinal character of the graduality of fuzzy sets, and due to eliminating all fuzzy connectives, the dominance-based rough approximations of fuzzy sets give a new insight into both rough sets and fuzzy sets and enable further generalizations of both of them. The recently proposed DRSA for fuzzy case-based reasoning is an example of this capacity. This article also exhibits some important merits of DRSA within granular computing, which can be summarized as follows:

- DRSA extends the paradigm of granular computing to problems involving ordered data.
- It specifies a syntax and modality of information granules, defined by means of dominance-based constraints, which are appropriate for dealing with ordered data.
- It provides a methodology for dealing with this type of information granules, which results in a theory of computing with words and reasoning about ordered data.
- It is supported by a robust and meaningful algebraic model, bipolar Pawlak-Brouwer-Zadeh lattice, and this ensures the solidity of the obtained results.

Future Directions

The granular computing with ordered data is a very general problem, because also other

modalities of information constraints, such as veristic, possibilistic and probabilistic modalities, have to deal with ordered value sets (with qualifiers relative to grades of truth, possibility, and probability). For this reason, granular computing with ordered data based on DRSA is a very promising field of research. There is a great potential in both theoretical investigations, such as extension of DRSA to other algebraic models than the bipolar Pawlak-Brouwer-Zadeh lattice (for a general survey of algebraic structures for rough set theory, see Chap. 12 in Polkowski (2002)), and practical applications, such as customization of DRSA to case-based reasoning in specific domains, like medical diagnosis or cybersecurity.

Bibliography

- Błaszczyński J, Greco S, Słowiński R, Szeląg M (2009) Monotonic variable consistency rough set approaches. *Int J Approx Reason* 50:979–999
- Cattaneo G (1997) Generalized rough sets (Preclusivity fuzzy-intuitionistic (BZ) lattices). *Stud Logica* 58:47–77
- Cattaneo G, Ciucci D (2004) Algebraic structures for rough sets. In: Peters JF (ed) *Transaction on rough sets II*, LNCS, vol 3135. Springer, Berlin, pp 208–252
- Cattaneo G, Nisticò G (1989) Brouwer-Zadeh Poset and three-valued Łukasiewicz Posets. *Fuzzy Sets Syst* 33:165–190
- Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. *Int J Gen Syst* 17:191–208
- Dubois D, Prade H (1992a) Gradual inference rules in approximate reasoning. *Inf Sci* 61:103–122
- Dubois D, Prade H (1992b) Putting rough sets and fuzzy sets together. In: Słowiński R (ed) *Intelligent decision support – handbook of applications and advances of the rough sets theory*. Kluwer, Dordrecht, pp 203–232
- Dubois D, Prade H, Esteva F, Garcia P, Godo L, Lopez de Mantara R (1998) Fuzzy set modelling in case-based reasoning. *Int J Intell Syst* 13:345–373
- Dubois D, Grzymala-Busse J, Inuiguchi M, Polkowski L (eds) (2004) *Transactions on rough sets II: rough sets and fuzzy sets*, LNCS, vol 3135. Springer, Berlin
- Dyer J (2005) MAUT – multiattribute utility theory, chapter 7. In: Figueira J, Greco S, Ehrgott M (eds) *Multiple criteria decision analysis: state of the art surveys*. Springer, Berlin, pp 266–294
- Figueira J, Greco S, Ehrgott M (eds) (2005) *Multiple criteria decision analysis: state of the art surveys*. Springer, Berlin
- Fodor J, Roubens M (1994) *Fuzzy preference modelling and multicriteria decision support*. Kluwer, Dordrecht
- Fortemps P, Greco S, Słowiński R (2008) Multicriteria decision support using rules that represent rough-graded preference relations. *Eur J Oper Res* 188:206–223
- Gilboa I, Schmeidler D (2001) *A theory of case-based decisions*. Cambridge University Press, Cambridge
- Ginsburg S, Hull R (1983) Order dependency in the relational model. *Theor Comput Sci* 26:149–195
- Greco S, Matarazzo B, Słowiński R (1999) The use of rough sets and fuzzy sets in MCDM, chapter 14. In: Gal T, Stewart T, Hanne T (eds) *Advances in multiple criteria decision making*. Kluwer Academic Publishers, Boston, pp 14.1–14.59
- Greco S, Matarazzo B, Słowiński R (2000a) Rough set processing of vague information using fuzzy similarity relations. In: Calude C, Paun G (eds) *From finite to infinite*. Springer, Berlin, pp 149–173
- Greco S, Matarazzo B, Słowiński R (2000b) A fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. In: Fodor J, De Baets B, Perny P (eds) *Preferences and decisions under incomplete information*. Physica-Verlag, Heidelberg, pp 131–154
- Greco S, Matarazzo B, Słowiński R (2001a) Rough sets theory for multicriteria decision analysis. *Eur J Oper Res* 129:1–47
- Greco S, Matarazzo B, Słowiński R (2001b) Rough set approach to decisions under risk. In: Ziarko W, Yao Y (eds) *Rough sets and current trends in computing*, LNAI, vol 2005. Springer, Berlin, pp 160–169
- Greco S, Matarazzo B, Słowiński R, Stefanowski J (2001c) Variable consistency model of dominance-based rough set approach. In: Ziarko W, Yao Y (eds) *Rough sets and current trends in computing*, LNAI, vol 2005. Springer, Berlin, pp 170–181
- Greco S, Inuiguchi M, Słowiński R (2002a) Dominance-based rough set approach using possibility and necessity measures. In: Alpignini JJ, Peters JF, Skowron A, Zhong N (eds) *Rough sets and current trends in computing*, LNAI, vol 2475. Springer, Berlin, pp 85–92
- Greco S, Matarazzo B, Słowiński R (2002b) Preference representation by means of conjoint measurement and decision rule model. In: Bouyssou D, Jacquet-Lagrange E, Perny P, Słowiński R, Vanderpooten D, Vincke P (eds) *Aiding decisions with multiple criteria – essays in honor of Bernard Roy*. Kluwer, Dordrecht, pp 263–313
- Greco S, Prędko B, Słowiński R (2002c) Searching for an equivalence between decision rules and concordance-discordance preference model in multicriteria choice problems. *Control Cybern* 31:921–935
- Greco S, Inuiguchi M, Słowiński R (2004a) A new proposal for rough fuzzy approximations and decision rule representation. In: Dubois D, Grzymala-Busse J, Inuiguchi M, Polkowski L (eds) *Transactions on rough sets II: rough sets and fuzzy sets*, LNCS, vol 3135. Springer, Berlin, pp 156–164
- Greco S, Matarazzo B, Słowiński R (2004b) Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *Eur J Oper Res* 158:271–292

- Greco S, Matarazzo B, Słowiński R (2004c) Dominance-based rough set approach to knowledge discovery (I) – general perspective, chapter 20. In: Zhong N, Liu J (eds) *Intelligent technologies for information analysis*. Springer, Berlin, pp 513–552
- Greco S, Matarazzo B, Słowiński R (2004d) Dominance-based rough set approach to knowledge discovery (II) – extensions and applications, chapter 21. In: Zhong N, Liu J (eds) *Intelligent technologies for information analysis*. Springer, Berlin, pp 553–612
- Greco S, Matarazzo B, Słowiński R (2005) Generalizing rough set theory through dominance-based rough set approach. In: Ślęzak D, Yao J, Peters J, Ziarko W, Hu X (eds) *Rough sets, fuzzy sets, data mining, and granular computing*, LNAI, vol 3642. Springer, Berlin, pp 1–11
- Greco S, Inuiguchi M, Słowiński R (2006a) Fuzzy rough sets and multiple-premise gradual decision rules. *Int J Approx Reason* 41:179–211
- Greco S, Matarazzo B, Słowiński R (2006b) Dominance-based rough set approach to case-based reasoning. In: Torra V, Narukawa Y, Valls A, Domingo-Ferrer J (eds) *Modelling decisions for artificial intelligence*, LNAI, vol 3885. Springer, Berlin, pp 7–18
- Greco S, Matarazzo B, Słowiński R (2007) Dominance-based rough set approach as a proper way of handling graduality in rough set theory. In: Peters JF (ed) *Transactions on rough sets VII*, LNAI, vol 4400. Springer, Berlin, pp 36–52
- Greco S, Matarazzo B, Słowiński R (2012a) Distinguishing vagueness from ambiguity by means of Pawlak-Brouwer-Zadeh lattices. In: Greco S et al (eds) *International conference on information processing and management of uncertainty in knowledge-based systems*. Springer, Berlin, pp 624–632
- Greco S, Matarazzo B, Słowiński R (2012b) The bipolar complemented de Morgan Brouwer-Zadeh distributive lattice as an algebraic structure for the dominance-based rough set approach. *Fund Inform* 115:25–56
- Greco S, Matarazzo B, Słowiński R (2016) Decision rule approach, chapter 13. In: Greco S, Ehrgott M, Figueira J (eds) *Multiple criteria decision analysis: state of the art surveys*. Springer, Berlin, pp 497–552
- Greco S, Matarazzo B, Słowiński R (2017) Distinguishing Vagueness from ambiguity in dominance-based rough set approach by means of a bipolar Pawlak-Brouwer-Zadeh lattice. In: Polkowski L et al (eds) *International joint conference on rough sets, IJCRS 2017, Part II*, LNAI 10314. Springer, Berlin, pp 81–93
- Hume D (1748) *An enquiry concerning human understanding*. Clarendon Press, Oxford
- Klement EP, Mesiar R, Pap E (2000) *Triangular norms*. Kluwer Academic Publishers, Dordrecht
- Kolodner J (1993) *Case-based reasoning*. Morgan Kaufmann, San Mateo
- Leake DB (1996) CBR in context: the present and future. In: Leake D (ed) *Case-based reasoning: experiences, lessons, and future directions*. AAAI Press/MIT Press, Menlo Park, pp 1–30
- Lin TY (1988) Neighborhood systems and relational databases. In: *Proceedings of the ACM conference on computer science*, p 725
- Lin TY (1989) Neighborhood systems and approximation in database and knowledge base systems. In: *Proceedings of the fourth international symposium on methodologies of intelligent systems, poster session*, 12–15 Oct 1989, pp 75–86
- Lin TY (1992) Topological and fuzzy rough sets. In: Słowiński R (ed) *Intelligent decision support – handbook of application and advances of the rough sets theory*. Kluwer Academic Publishers, Dordrecht, pp 287–304
- Lin TY (1997) Granular computing. In: *Announcement of the BISC special interest group on granular computing*
- Lin TY (1998a) Granular computing on binary relations I: data mining and neighborhood systems. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*. Physica-Verlag, Heidelberg, pp 107–121
- Lin TY (1998b) Granular computing on binary relations II: rough set representations and belief functions. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*. Physica-Verlag, Heidelberg, pp 121–140
- Loemker L (ed and trans), Leibniz GW (1969) *Philosophical papers and letters*, 2nd edn. D. Reidel, Dordrecht
- Nakamura JM (1991) Gao, a logic for fuzzy data analysis. *Fuzzy Sets Syst* 39:127–132
- Pal SK, Skowron A (eds) (1999) *Rough-fuzzy hybridization: a new trends in decision making*. Springer, Singapore
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11:341–356
- Pawlak Z (1991) *Rough sets*. Kluwer, Dordrecht
- Pawlak Z (2001) *Rough set theory*. *Kunstliche Intelligenz* 3:38–39
- Peters JF, Skowron A, Dubois D, Grzymala-Busse J, Inuiguchi M, Polkowski L (eds) (2005) *Rough sets and fuzzy sets, transaction on rough sets II*. Springer, Berlin
- Polkowski L (2002) *Rough sets: mathematical foundations*. Physica-Verlag, Heidelberg
- Radzikowska M, Kerre EE (2002) A comparative study of fuzzy rough sets. *Fuzzy Sets Syst* 126:137–155
- Słowiński R, Yao Y (eds) (2015) *Rough sets*. In: Kacprzyk J, Pedrycz W (eds) *Part C of the handbook of computational intelligence*. Springer, Berlin, pp 329–451
- Słowiński R, Greco S, Matarazzo B (2002a) Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. *Control Cybern* 31:1005–1035
- Słowiński R, Greco S, Matarazzo B (2002b) Mining decision-rule preference model from rough approximation of preference relation. In: *Proceedings of 26th IEEE annual international conference on computer software & applications (COMPSAC 2002)*, Oxford, pp 1129–1134
- Słowiński R, Greco S, Matarazzo B (2014) Rough set based decision support, chapter 19. In: Burke EK, Kendall G (eds) *Search methodologies: introductory*

- tutorials in optimization and decision support techniques, 2nd edn. Springer, New York, pp 557–609
- Stewart T (2005) Dealing with uncertainties in MCDA, chapter 11. In: Figueira J, Greco S, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, Berlin, pp 445–470
- Szeląg M, Greco S, Błaszczyński J, Słowiński R (2011) Case-based reasoning using dominance-based decision rules. In: Yao JT et al (eds) RSKT 2011, LNCS, vol 6954. Springer, Heidelberg, pp 404–413
- Szeląg M, Greco S, Słowiński R (2016) Similarity-based classification with dominance-based decision rules. In: Flores V et al (eds) IJCRS 2016, LNAI, vol 9920. Springer, Berlin, pp 355–364
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1979) Fuzzy and information granularity. In: Gupta M, Ragade RK, Yager RR (eds) Advances in fuzzy set theory and applications. North-Holland Publishing Company, Amsterdam, pp 3–18
- Zadeh LA (1996) Key roles of information granulation and fuzzy logic in human reasoning, concept formulation and computing with words. In: Proceedings of the 5th IEEE international conference on fuzzy systems, p 1
- Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 90:111–127
- Zadeh LA (1999) From computing with numbers to computing with words – from manipulation of measurements to manipulation of perception. *IEEE Trans Circuits Syst I Fund Theory Appl* 45:105–119
- Ziarko W (1993) Variable precision rough sets model. *J Comput Syst Sci* 46:39–59
- Ziarko W (1998) Rough sets as a methodology for data mining. In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery, vol 1. Physica-Verlag, Heidelberg, pp 554–576

Granular Computing, Information Models for

Steven A. Demurjian¹,
Alberto De la Rosa Algarin² and
Rishi Knath Saripalle³

¹Department of Computer Science and
Engineering, University of Connecticut, Storrs,
CT, USA

²Cynnovative, LLC, Arlington, VA, USA

³University of Connecticut, Storrs, CT, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Candidate Information Models
Suitability of Information Models for Granular
Computing
Future Directions
Bibliography

Glossary

Attribute-based data model (ABDM) ABDM is a data (information) model that organizes data (information) in attribute-value pairs, providing a means for both definition and access.

Data (information) model A data (information) model is used in database design to capture the structure or schema of the data. All data (information) that is entered into a database must conform to the definitions of the data (information) model.

Data (information) table A set of similar granules used in granular computing are collected into a data (information) table in order to allow them to be conceptualized and reasoned on in a formal manner.

Extensible markup language (XML) XML is a data (information) model that is a standard for exchanging and sharing data (information) across the Internet, among databases, and so on.

Functional data model (FDM) FDM is a semantic data (information) model to represent data (information) as it appears in the “real world” with a functional/logic basis.

Granular computing A discipline of information theory/computer science that uses a formal theory to reason about and analyze data (information) in granules.

Granule A piece of data (information) of varied size and complexity that is used to represent data (information) as it occurs in some “real-world” context.

Relational data model (RDM) RDM is a dominant data (information) model in commercial and open source database management systems with a basis in set theory.

Resource description framework (RDF) RDF is a meta-model to represent web data using expressions in the form of subject-predicate-object expressions.

Web ontology language (OWL) OWL is an extension of RDF and is a knowledge representation framework to capture domain knowledge in order to attach semantics to the represented information.

Definition of the Subject

Granular computing (GrC) is an emerging discipline of information theory that strives to allow reasoning and analysis based on varying levels of information granularity (from fine to coarse). In GrC, the entire information universe can be organized based on many different criteria, allowing the information to be abstracted, aggregated, classified, generalized, and so on, based on various characteristics (e. g., data similarity, operational usage, etc.). As a result, GrC relies on information models to describe the universe, the elements of

the universe, and the composition of each element. In this entry, alternative candidate information models for GrC are explored, including the attribute-based data model, the relational data model, the functional data model, the extensible markup language, the resource description framework, and the web ontology language. This includes both a description of these models and an analysis of the suitability in support of GrC.

Introduction

In information theory, one approach to support the reasoning and analysis of information based on varied levels of conceptualization is the emerging discipline of *granular computing* (GrC) (Lin 1997, 2005, 2006; Zadeh 1998), a term jointly coined by Lin and Zadeh and having a basis in research conducted by Lin on neighborhoods (e.g., granules) in databases (Lin 1988) and computer security (Lin 1989). Information granulation is an approach that partitions the entire information universe into *granules* (Lin 1998a, b; Zadeh 1996, 1997). This partitioning can be based on many different criteria: aggregating objects that demonstrate similarity of features or usage by classification (e.g., defining a relation such as a Course table in a university application); abstracting away details via generalization (e.g., creating an object-oriented class Person that would be the root of a hierarchy); distinguishing among different characteristics of objects by specialization (e.g., Students or Faculty subclasses of Person); summarizing the characteristics of similar objects into a coarser conceptualization (e.g., collecting multiple Computer Science Faculty instances into a single object); and so on. In all of these situations, the information universe is described in its entirety, and from that initial characterization, it is possible to define elements and components of elements.

In support of GrC, one of the key initial considerations is to define the information universe via an appropriate data representation. For GrC, this data representation has most frequently been based on data tables as formally defined for rough sets (Pawlak 1991), where a *data table* has many

different equivalent nomenclatures, including a knowledge representation system, an attribute-value system, and an information table. For a data table, columns are labeled as attributes (of the objects), and rows are labeled as instances (of the objects). Formally, the data table is characterized as a pair that contains a finite non-empty universe U and a finite non-empty set of primitive attributes A . Additionally, for each a in A , there is a non-empty set of values and a function that maps the universe U to these values. Collectively, these data tables based on rough sets have been used as the basis to define foundational models for granular computing (Lin 2005, 2006; Skowron and Stepaniuk 2001) and to achieve privacy protection for medical data by partitioning attributes into identifying, easily known, and sensitive categories (Wang et al. 2004).

The main objective of this entry is to explore alternative candidate information models for GrC to support the data table (and dependencies that may exist both within a table and across multiple data tables). Specifically, a variety of models that are old and new are explored, including the attribute-based data model (Hsiao and Harary 1970), the relational data model (Codd 1970), the functional data model (Shipman 1981), the extensible markup language (World Wide Web Consortium), the resource description framework (Powers 2003), and the web ontology language (Allemang and Hendler 2011). The *attribute-based data model* (ABDM), proposed in the early 1970, was touted for its ability to formally define “real-world” records of information, to support their retrieval, and has been shown capable of capturing data from relational, hierarchical, network, and functional data models (Demurjian and Hsiao 1987, 1988). The *relational data model* (RDM) proposed in mid-1970 along with the current SQL 1992 (SQL2) has evolved into the dominant database model for commercial and open source database systems, with extensions to support object-oriented and other features underway (SQL3). The *functional data model* (FDM) and its associated Daplex programming language, proposed in 1981, offer unique programming-like features, and its functional formal underpinnings can provide a means to reason about data and

constraints and, moreover, to capture semantics. The *extensible markup language (XML)* has emerged as an information-representation standard format, allowing information to be modeled and more easily exchanged among programs, websites, databases, etc. The *resource description framework (RDF)* leverages the structure of XML by annotating data and its structure with semantics using expressions in the form of subject-predicate-object triples. Finally, the *web ontology language (OWL)* is built on top of RDF to take advantage of RDF triples and provides richer semantics with greater expressive power for defining complex ontologies. All of these models offer different capabilities in support of data tables and GrC.

The remainder of this entry contains two sections and a conclusion. In section “[Candidate Information Models](#),” the six information models, ADBM, RDM, FDM, XML, RDF, and OWL, are examined. Using this as a basis, in section “[Suitability of Information Models for Granular Computing](#),” the suitability of the six information models in support of GrC is investigated by considering them against three different criteria. Finally, section “[Future Directions](#)” offers concluding remarks.

Candidate Information Models

In this section, background on the six information models to be considered for their suitability for GrC are reviewed, namely, the attribute-based data model (ABDM) (Hsiao and Harary 1970), the relational data model (RDM) (Codd 1970), the functional data model (FDM) (Shipman 1981), the extensible markup language (XML) ([World Wide Web Consortium](#)), the resource description framework (RDF) (Powers 2003), and the web ontology language (OWL) (Allemang and Hendler 2011). To serve as a common context for the discussion, a university database is utilized as an example. In the university database, faculty are tracked by name, identifier, office phone, and department (e.g., computer science, mathematics, English, etc.), and students are tracked by name, identifier, grade point average, and campus address (e.g., dormitory). For each course in the

catalog, there is a unique course number (e.g., a combination of department and number such as CS123, MATH233, etc.), the course title and description, and the list of prerequisites for the course. The courses offered are also tracked, by course number/section number pairs (e.g., courses may be offered in multiple sections), the FacultyID teaching each offering (by faculty identifier), and the term (e.g., semester, quarter, etc.) of the offering (e.g., Fall2007, Spring2008, Summer2008, etc.). Likewise, the students enrolled by the course number/section number pair for each term are tracked.

The Relational Data Model (RDM)

Relational data is organized into tuples of *relations*. A database is a collection of relations. The attributes of a relation are distinct. The tuples of a relation have the property that no two tuples are identical. In addition, one or more attributes of the relation may be defined as the *primary key* of the relation; a relation can have multiple candidate keys, one of which is chosen as the primary key. The primary key of the relation is used to uniquely identify the tuples of the relation. To establish dependencies across two or more relations, a *foreign key* can be defined consisting of one or more attributes of one relation that reference the primary key of another relation. Foreign keys are employed to establish referential integrity constraints across two relations where the foreign key of one relation references the primary key of another relation (with self-reference allowed). Non-primary key attributes are allowed to have null values; this includes a foreign key. A definition of the university database in tabular (or relational) form is as below:

```
Student(Name, studentID*, GPA, CampusAddress)
Faculty(Name, FacultyID*, Ophone, Department)
Courses(Course#, Title, Department,
                                             PCourse#)
OfferedCourses(Course#, Section#,
                                             FacultyID*, Term*)
EnrolledStudents(Course#, Section#,
                                             StudentID*, Term*)
```

Briefly, let us describe the data relationships of this database. First, each student of the Student

relation is uniquely identified by a StudentID (primary key), has a Name, a grade point average (GPA), and a CampusAddress. Next, each faculty member of the Faculty relation also has a unique FacultyID (primary key), a Name, an office phone number, and a Department affiliation. Third, each course of the Courses relation (catalog) is uniquely identified by a Course# (primary key), has a Title and a Description, and may have zero or more prerequisite courses (PCourse#-foreign key). Fourth, the OfferedCourses relation tracks the courses offered by sections taught by faculty each term, with these four attributes forming a compound primary key. Fifth, the EnrolledStudents relation tracks the students enrolled in sections of courses for each term, again by a compound primary key. In terms of referential integrity, the foreign key PCourse# self-references the relation Course; note that a course without a prerequisite will have a null foreign key.

The Attribute-Based Data Model (ABDM)

In ABDM, the initial step in defining a database is the specification of a collection of the attributes that are in the database. The *database records* in ABDM consist of sets of attribute-value pairs. An *attribute-value pair* is a member of the Cartesian product of the attribute name and the value domain of the attribute. As an example, <Department, Computer Science> is an attribute-value pair having Computer Science as the value for the Department attribute. Database records have the property that for a given set of attribute-value pairs, no two pairs in the set have the same attribute name. In ABDM, data is considered in the following constructs: database, file, record, attribute-value pair (keyword)/attribute-value range, record body, directory, directory keyword, and non-directory keyword. Informally, a *database* consists of a collection of files. Each *file* contains a group of records which are characterized by a unique set of keywords. A *record* is composed of two parts. The first part is a collection of *attribute-value pairs* (*keywords*). The second part of the record, which is optional, is for unformatted textual information and is referred to as the *record body*. A record contains at most one attribute-value pair for each

attribute defined in the database. An example of a record equivalent to the Faculty relation is:

```
(<File, Faculty>, <Record Number, 123>,
<Name, John Smith>, <FacultyID, 12121212>,
<ophone, 5551111>, <Department, Computer
                        Science>,
{This is an assistant professor up for
tenure in 2012})
```

The angle brackets, <, >, enclose an attribute-value pair, i.e., keyword. The curly brackets, {, }, include the record body. The record is enclosed in parentheses. The first attribute-value pair of all records of a file, by convention, is the same. In particular, the attribute is File, and the value is the file name (akin to a relation name in RDM). Similar records for Student, Courses, CourseOfferings, and EnrolledStudents can also be defined; referential integrity is achieved by using the unique record numbers that are given to every instance.

Finally, in ABDM, the indexing criteria for a given database is also a critical part of the definition of the database. In particular, for indexing purposes, there are two different types of attribute-value pairs in a record (or a file), with all of the indexing data maintained in the *directory* of the database. Certain attribute-value pairs of a record (or a file) are called the *directory keywords* of the record (file), because either the attribute-value pairs or their attribute-value ranges are kept in a directory for identifying the records (files). Those attribute-value pairs which are not kept in the directory are called *non-directory keywords*. In the example record above, the File, Record Number, and FacultyID attributes would be directory keywords, while the other attributes would be non-directory keywords. The identification of database records are by either directory or non-directory keywords. When directory keywords are used, the search space is clearly defined using the indexing criteria. When non-directory keywords are used, the entire file of records must be searched. Note that null values are not allowed for directory attributes.

The Functional Data Model (FDM)

An *entity* in a functional database is always associated with a collection of distinct functions that

can be applied to the entity to return either individual data values or one or more objects. The term *object* is used to refer to the actual data values (values for the functions) for an entity. Thus, an object can be considered as an instantiation (occurrence, in the earlier terminology) of the entity. An object is analogous to a tuple in the RDM or a record in ABDM. The *functions* of an entity are applied to the entity to return a particular value that is associated with that entity, i.e., to return a portion of the object. There are two types of functions, scalar-valued functions and entity-valued functions. Each type of function may be either single-valued (returning one value) or set-valued (returning zero or more values). *Scalar-valued functions* return one or more typical database values (i.e., string, integer, and float). *Entity-valued functions* return one or more entity objects as their values.

Additionally, in the FDM, relationships between the entities may be defined that lead to one or more generalization hierarchies for the database. Generalization (Smith and Smith 1977) is an abstraction technique that is used to organize commonalities from multiple entities into a single entity (e.g., the name and identifier for students and faculty can be organized in a person entity). Because of the presence of generalization hierarchies, there are two different categories of entities, entity subtypes and entity supertypes. An *entity subtype* exists at the inner and leaf nodes of a generalization hierarchy and inherits all of the characteristics of its ancestors, i.e., inherits all of the functions of its ancestors. An *entity supertype*, or, more simply, an *entity type*, is at the root of a generalization hierarchy and has the property that no two instantiations (occurrences) of the entity type return the same values for all of the entity's functions. Additionally, one or more single-valued, scalar-valued functions in an entity type may be defined to be the key of the entity. Functional data is organized into *generalization hierarchies* of *entities*. A database is a collection of entities organized into generalization hierarchies. References among entities are accomplished when a function is entity valued, with self-referencing possible. In the following, the definition of the FDM version of the university database using the Daplex data definition language is presented.

```

DATABASE University IS

TYPE      Perso;
SUBTYPE   Student;
SUBTYPE   Faculty;
TYPE      Courses;
TYPE      CoursesOffered;

TYPE Person IS
    Name      : STRING(1..30);
    Identifier : STRING(1..9);
END ENTITY;

TYPE Courses IS
    Course#    : STRING(1..8);
    Title      : STRING(1..20);
    Descrip    : STRING(1..100);
    PCourse#   : SET OF Courses;
END ENTITY;

TYPE CoursesOffered IS
    Course#    : STRING(1..8);
    Section#   : INTEGER;
    TERM       : STRING(1..10);
END ENTITY;

SUBTYPE Student IS Person
    StTakes : SET OF CoursesOffered;
    SPA     : FLOAT;
    campusAddress : STRING(1..100);
END ENTITY;

SUBTYPE Faculty IS Person
    Teaches : SET OF CoursesOffered;
    Ophone  : STRING(1..7);
    Department : STRING(1..30);
END ENTITY;

UNIQUE Course# WITHIN Courses;
UNIQUE Section# WITHIN Formats;
UNIQUE Identifier WITHIN Person;
UNIQUE Course#, Section# Term WITHIN
                                CoursesOffered;

END UNIVERSITY;

```

The FDM version of the university database has subtle differences from its RDM counterpart. Note that Student and Faculty in the RDM version have been reorganized into a generalization hierarchy with Person (root entity type with commonalities) and Student and Faculty (children or entity subtypes). Courses is similar, but the foreign key PCourse# has been replaced by an entity-valued function. The other two differences are the following: CoursesOffered no longer includes the faculty identifier and has been replaced by a Teaches entity-valued function in Faculty; and EnrolledStudents has been replaced by a Takes

entity-valued function in Student. The end of the definition of the database includes the various uniqueness constraints (akin to primary keys).

The Extensible Markup Language (XML)

The extensible markup language (XML) ([World Wide Web Consortium](#)) has emerged as a standard for information modeling and exchange for web-based applications, database interoperability, common software tool formats, patient record data, etc. XML allows information content to be hierarchically organized and tagged to highlight important/relevant content. The *tags* capture not only the content but can be leveraged to represent the meaning of the information (semantics). In a web-based setting, XML is the successor to HTML to allow information content to be hierarchically organized and tagged to highlight important and relevant content. The tags can be exploited to capture both information content and the meaning of the information (semantics). In addition, XML allows the definition of templates to capture the known structure information for an application by the creating of XML schema files called Document Type Definitions (DTDs). The resulting XML document instances that are created contain both information content (data) and semantic notations (tags) that indicate the meaning of the information.

The extensible markup language (XML) ([World Wide Web Consortium](#)) provides a flexible means to store and transmit data between different information systems and platforms and has emerged as a dominant means for interoperability on the World Wide Web (as a successor to HTML) and as a standard for information format and exchange (e.g., the OpenDocument project for office applications). Both HTML and XML use tags to identify data/elements. However, while the HTML tags are predefined and specify the way the data within the tags is to be displayed, the XML tags are user-defined and can be employed to identify the data structure in a hierarchical fashion. For instance, in HTML, `< b >` or `< / b >` are all predefined tags to display the data/characters between them in a bold font. In XML, you will not have those kinds of predefined tags. All of the tags are defined by the users to indicate the structure of the data elements. An XML document for faculty data would be as simple as:

```
<?xml version='1.0' encoding='ISO-8859-1' ?>
<Faculty>
  <Name> John Smith </Name>
  <FacultyID> 12121212 </FacultyID>
  <Ophone> 5551111 </Ophone>
  <Department> Computer Science </Department>
</Faculty>
```

In the XML fragment, the tags `<Faculty>`, `<Name>`, `<FacultyID>`, etc., are defined, based on the application and/or modeling requirements. These tags capture data content (with attribute names) as well as data dependencies; they do not provide any clues on formatting or displaying the data elements between them. In addition to these structural capabilities of an XML file, each of the elements can be further quantified with tags to track semantic aspects of the data. For example, the Name could be augmented with a tag that indicates whether it is a married or maiden name, and the Department could be tagged with a title (e.g., Professor, Associate Professor, Head, etc.). If the XML structure contained numerical data (e.g., GPA), then the additional tags may denote the scale of the data (e.g., out of 4.0, 100, etc.).

The overall appearance and structure of an XML file can also be defined and validated by an XML schema file. An XML schema file defines the tags or attributes that can be used and where they appear in an XML file. There are two types of XML schema files: the Document Type Definition (DTD) and the XML Schema (a more sophisticated construct). A sample XML Schema for the previous XML fragment can be defined as:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xs:schema xmlns:xs=
  "http://www.w3.org/2001/XMLSchema">

  <xs:element name="Faculty">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="Name"
          type="xs:string"/>
        <xs:element unique name="FacultyID"
          type="xs:string"/>
        <xs:element name="Ophone"

        <xs:element name="Department"
          type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
```

This sample XML Schema file defines that in the corresponding XML file, the Faculty element is a complex type (akin to a relation in RDM, file in ABDM, and entity type in FDM) and must have one Name, one FacultyID, one Ophone, and one Department element. Uniqueness is achieved via the “unique” keyword attached to the FacultyID element. With a XML Schema, an XML parser can be used to validate whether the elements within a XML file are valid or not with respect to the XML Schema. Conceptually, a XML Schema is analogous to type declaration for a RDM relation, a ABDM file, or a FDM entity type (subtype).

The Resource Description Framework (RDF)

The *resource description framework (RDF)* (Powers 2003) is a data model that revolves around the idea of conceptual modeling via resource statements. Unlike traditional data models, where an object is described by the attributes and their values, in RDF, these expressions are known as *triples*, divided into a *subject*, a *predicate*, and an *object*. The *subject* represents the resource (e.g., Faculty), the *predicate* represents the attributes or other descriptive aspects of the resource (e.g., Name, FacultyID, Ophone, Department), while at the same time denoting the relationship of the resource with the *object* (e.g., the triple <John Smith, 12121212, 5551111, Computer Science>).

RDF builds on top of XML (with XML being one of the possible serialization formats of RDF) and is one of W3C’s Semantic Web components. At the specification level, the RDF Schema includes definitions for *classes* (model level) and *properties* (instance level), including classes for statements, properties, resources, etc. and properties for subject items, objects (of subjects), predicates, etc. Following the Faculty example of this entry, an RDF for the Faculty resource with Name, FacultyId, Ophone, and Department predicates and respective objects (values) would appear as below via XML serialization:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/
1999/02/22-rdf-syntax-ns#"

```

```
xmlns:dc="http://www.university.
example/#Faculty">
<rdf:Description rdf:about="http://
www.university.example/#Faculty">
  <dc:Name>John Smith</cd:Name>
  <dc:FacultyID>12121212</cd:
FacultyID>
  <dc:Ophone>5551111</cd:Ophone>
  <dc:Department>Computer Science</cd:
Department>
</rdf:Description>
</rdf:RDF>
```

While RDF is typically serialized utilizing the XML format, the Notation 3 (N3) format is also utilized for RDF serialization in common cases. As an example of the N3-RDF serialization, the previous Faculty example changes to:

```
@prefix dc: <http://www.university.
example/#Faculty>.
<http://www.university.example/
#Faculty>
dc:Name "John Smith"
dc:FacultyID "12121212"
dc:Ophone "5551111"
dc:Department "Computer Science"
```

Various query and interface languages are available to obtain information from RDF graphs. The most prominent is SPARQL, a query language with similar traits to SQL. Lastly, thanks to its capabilities, RDF is utilized throughout a series of domains for different purposes. For example, it is utilized to embed license information, for publishing updates in a similar fashion to RSS feeds, etc.

The Web Ontology Language (OWL)

The Web Ontology Language (OWL) (Allemang and Hendler 2011) has established itself as a standard for developing ontologies which capture semantics for a domain. The fundamental modeling constructs of the OWL framework are based on *SHOIN(D)* description language, which is built on top of *Attributive Concept Language with Complements (ALC)* description logic. The OWL framework is built on top of RDF and XML, to exploit the *Triple* knowledge statement format of RDF and the tag based representation of

information from XML. The knowledge semantics captured in the ontology can be attached to digital data such as electronic health records, biological and chemical information, XML instances, database entities, etc. represented using various modeling standards such as XML, HTML, RDM, etc. The Faculty and Course concepts are represented in OWL as:

```
<owl:Class rdf:about="http://www.owl-ontologies.com/GrC.owl\#Faculty">
<owl:Class rdf:about="http://www.owl-ontologies.com/GrC.owl\#Course">
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#FacultyID">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Faculty"/>
  <rdfs:range rdf:resource="xsd;integer"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#Ophone">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Faculty"/>
  <rdfs:range rdf:resource="xsd;integer"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#Department">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Faculty"/>
  <rdfs:range rdf:resource="xsd;string"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#CourseID">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Course"/>
  <rdfs:range rdf:resource="xsd;integer"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#Title">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Course"/>
  <rdfs:range rdf:resource="xsd;integer"/>
```

```
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:
about="http://www.owl-ontologies.com/GrC.owl\#Description">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/GrC.owl\#Course"/>
  <rdfs:range rdf:resource="xsd;string"/>
</owl:DatatypeProperty>
<owl:ObjectProperty rdf:about="http://www.owl-ontologies.com/GrC.owl\#teaches">
  <rdfs:domain rdf:resource="http://www.owl-ontologies.com/Cardiology.owl\#Faculty"/>
  <rdfs:range rdf:resource="http://www.owl-ontologies.com/Cardiology.owl\#Course"/>
</owl:ObjectProperty>
```

The OWL code as given above has the concepts Faculty and Course are defined as classes using the OWL standard tag `<owl:Class>` and described using properties FacultyID, Ophone, Department, CourseID, Title, and Description, which are defined to hold datatype values represented using the OWL `<owl:DatatypeProperty>` tag. The two concepts are related to each other with association *teaches* (read as Faculty teaches Course) and is represented using `<owl:ObjectProperty>`. This OWL structure can be instantiated to captures instance as below:

```
<owl:NamedIndividual rdf:
about="http://www.owl-ontologies.com/GrC.owl\#John_Smith">
  <rdf:type rdf:resource="http://www.owl-ontologies.com/Cardiology.owl\#Faculty"/>
  <FacultyID rdf:datatype="xsd;string">12121212</FacultyID>
  <Ophone rdf:datatype="xsd;string">5555511111</OPhone>
  <Department rdf:datatype="xsd;string">Computer Science</OPhone>
  <teaches rdf:resource="http://www.owl-ontologies.com/GrC.owl\#OWL_Introduction"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:
about="http://www.owl-ontologies.com/GrC.owl\#OWL_Introduction">
```



```

<rdf:type rdf:resource="http://
www.owl-ontologies.com/Cardiology.owl
\#Course"/>
<CourseID rdf:datatype="xsd;
string">CSE101</CourseID>
<Title rdf:datatype="xsd;
string">OWL Semantic Web</Title>
<Description rdf:datatype="xsd;
string">Introduction to OWL</
Description>
</owl:NamedIndividual>

```

The OWL instance “John Smith” which is of type Faculty and “OWL Introduction” which is of type Course are captured using `<owl:NamedIndividual>`, and the other respective property values are filled to capture the knowledge about these concepts. When the presented OWL sample is compared to RDF and XML, the information is represented as a knowledge statement (e.g., Faculty *teaches* Courses) using tag like structure, but the OWL framework allows ontology developers to add various description-logic-based semantic features such as disjointness, union, negation, transitive, inverse, etc., on OWL classes and properties, making the knowledge represented using OWL semantically richer than RDF and XML data models. The OWL graph structured can be queried using SPARQL query language similar to RDF.

Suitability of Information Models for Granular Computing

This section compares the six information models as given in section “[Candidate Information Models](#)” (ABDM, RDM, FDM, XML, RDF, and OWL) for their suitability in support of granular computing. To facilitate the comparison, criteria are utilized that are relevant for comparing data models (Demurjian and Hsiao 1987, 1988), in conjunction with various concepts proposed in (Zadeh 1997). The criteria chosen are:

- **Formalism and Theory:** This criterion is used to assess the degree to which a formal model exists for the data model that is capable of supporting GrC.

- **Expressive Power:** This criterion is used to evaluate the ability of each data model to support the definition, usage, and analysis of granules.
- **Extensibility and Utility:** This criterion is used to determine the usability of the data model, from many different perspectives.

In the remainder of this section, these criteria are explored against the six models, as a means to assess their suitability for GrC.

Formalism and Theory

All of the data models given in section “[Candidate Information Models](#),” have a certain degree of formalism/theory upon which they are based. GrC relies on an information model that has a formal basis, in order to capture the required information within a data table (Pawlak 1991) and to employ a theory of rough sets and fuzzy logic for granular analysis. As a result, it is vital that an underlying data model is rich enough to support the formalisms necessary for GrC. To begin, the attribute-based data model (Hsiao and Harary 1970) was proposed by David K. Hsiao (founder of ACM Transactions on Database Systems) and Frank Harary (noted graph theorist) and has a formal basis that has been explored by other researchers in the 1970s (Rothnie Jr 1974), 1980s (Demurjian and Hsiao 1988), and 1990s (Lin 1992), with applicability in many different contexts including paging environments for OS, data-model transformations for federated databases, and polyinstantiation for security, respectively. Likewise, the relational data model (Codd 1970) has a rich history and formal basis using set theoretic concepts and is the dominant approach to date for GrC researchers (Demchenko et al. 2005; Lin 2005; Wang et al. 2004).

The functional data model (Shipman 1981) was proposed in an era in the early-to-mid 1980s when semantic data models (successors to the entity-relationship data model) were advocated to break away from relatively flat approaches (e.g., relational, hierarchical, and network data models and databases systems) to an approach that was semantically rich and able to model the

data as it occurs in the “real world.” In fact, in a workshop in 1993 later published as special journal issue (Gray et al. 1999), the overriding theme emphasized the functional model and its potential for a unifying paradigm, providing formalisms based on functional and logic programming that could be used to capture both relational data and object-oriented data in a formal way.

XML ([World Wide Web Consortium](#)) has emerged as a de facto standard across many different disciplines, being used in a wide variety of contexts (web semantics, patient health records/standards, database interoperability, information exchange format, etc.). XML is based on detailed specifications ([XML 1.0 Fourth Edition](#)) and has a grammar in Extended Backus-Naur Form. As a result, there are XML parsers that have been built for many purposes to parse XML files and XML Schemas (see section “[The Extensible Markup Language \(XML\)](#)”) again) in different ways. Such a formal-grammar basis means that XML can be utilized in a very rigorous manner to reason using context-free grammar concepts from automata theory.

RDF extends XML with semantic capabilities, and the data model is represented by a directed labeled graph. This formalism benefits RDF by providing an easy to read structure, as well as representing RDF data models that cannot be serialized with XML. This method of building an RDF resource also allows the combination of RDF triples (3-tuples) with other RDF triples. When merging two RDF graphs, blank nodes are not brought together since no automated method exists on deciding whether the nodes are equal. In that regard, an RDF graph is considered grounded if there exist no blank nodes. In the same manner that XML parsers exist to extract XML instances and schemas in different ways, RDF validators exist that return an instance of an RDF/XML serialization of a represented resource. This validation process returns more than one instance in those RDF graphs that are not grounded, but they would still be semantically equivalent. As stated in section “The Resource Description Framework (RDF),” SPARQL is utilized for querying the RDF graph structure to obtain information.

Lastly, the OWL framework is an emerging standard for ontology development, primarily employed for capturing domain semantics which can be later be attached electronically to represent information converting it into knowledge. The OWL knowledge framework is fundamentally based on SHOIN description logic with detailed semantics, which allows creating/editing of OWL-based ontologies using an application programmers interface (e.g., Jena API) and querying the ontology graph structure using the SPARQL query language. The semantic specifications based on description logic provide a formal framework for representation knowledge and inferring implicit knowledge using the represented knowledge through reasoning algorithms. In summary, it is clear that from a formal basis, each of these six data models can realize the information/data table needs of GrC, offering different strengths in their support.

Expressive Power

The expressive power needed to support information granules can be impacted by many factors; we choose three for our comparison:

- *Grouping capability* which involves the way that objects are grouped into granules (equivalence classes) and can be based on object relationships, similarity, proximity, semantics, etc. (Zadeh 1997)
- *Linguistic representation of granules* which assigns a linguistic value to each granule that can be accomplished using a naming convention or by providing sample representative objects (Zadeh 1997)
- *Constraint and type checking* which involves the ability to enforce intragranular and intergranular dependencies (constraint checking) while simultaneously insuring that the granule always adheres to the defined structure (type checking)

From an information model perspective, these three granulation factors are directly related to the features and characteristics of the data models as

presented in section “[Candidate Information Models](#)” and include aggregation (records in ABDM, relations in RDM, types and subtypes in FDM, XML Schemas in XML, graphs and subgraphs in RDF, and grouping similar objects or instances in OWL), generalization (inheritance in FDM and OWL), identity and uniqueness (directory attributes in ABDM, primary keys in RDM, uniqueness in FDM, unique in XML, URIs in RDF), and relationships (record references in ABDM, foreign keys in RDM, entity-valued functions in FDM, and hierarchical structure in XML, graph nodes and edges in RDF, and object properties in OWL).

Specifically, for the grouping capability factor, the assembly of objects into granules (equivalence classes) is aggregation, which is provided by all six data models. Aggregation is simply a result of the creation of the equivalence class, and the aggregation can be based on similarity, proximity, etc. The process of grouping objects into granules is part of GrC, but once the granulation has been completed, then any of these six models have sufficient aggregation to represent a granule. If granules are related to one another hierarchically (or the grouping occurs based on a hierarchical relationship), then either FDM or XML or OWL is appropriate for this resultant aggregation. RDF, while not suited primarily for hierarchical aggregation of equivalent structures, can be utilized as well. For the linguistic representation of granules factor, again, the same aggregation capability applies; all the models provide a naming convention, and the examples as given in section “[Candidate Information Models](#),” for ABDM, XML, and OWL illustrated data instances (sample representative objects).

Finally, for the constraints and type checking factor, each model offers different capabilities. For ABDM, uniqueness of records, record references across objects (instances), support for attribute-value pairs (used by many granulation researchers), and directory attributes (akin to keys), all provide the basis for constraint checking; the definition of the attributes (name and type) in each record provides type checking. For RDM, no duplicate tuples per relation, primary keys, foreign keys, and referential integrity

are relevant for constraint checking; a relational table definition with its attributes and data types provides type checking. For FDM, the types and subtypes provide explicit support for type checking (i.e., these types and subtypes are akin to programming languages types and subtypes in Ada), while uniqueness constraints preserve properties (constraint checking). For XML, the XML Schema is a template that an XML instance must follow (type checking) and, in its definition, provides the ability for dependencies (constraint checking); as such, for different granules, there would be different XML Schemas, and XML provides the ability to parse and map from one XML Schema to another via XSL Transformations (<http://www.w3.org/TR/xslt>) and XQuery (<http://www.w3.org/TR/xquery>). For RDF, the framework’s specification provides a vocabulary for constraints on the use of RDF properties and classes. Toward this end, the RDF schema provides mechanisms to define constraints, but it does not dictate their enforcement on the application level. The RDF schema specifies several constraints: property use with *rdfs:domain* and *rdfs:range*, no loops formed from the use of *rdfs:subPropertyOf* and *rdfs:subClassOf*, and any other constraints specified by the *rdfs:ConstraintResource* subclass. Lastly for OWL, the description logic framework provides features such as cardinality constraints on relationships, Boolean expressions such as union, disjoint, negation or complex logical expression on range values associated with relationships, primary key, etc. provide a formalism for constraint mechanism. XML is much more powerful than ABDM, RDM, FDM, and RDF, particularly in type checking; the use of XML can facilitate the ability to transform one granule to another while still preserving content and semantics. However, OWL is most powerful than XML, due to its expressive power provided by the description logic and exploiting it through inference algorithms.

Extensibility and Utility

The final criterion considered is extensibility and utility, which is focused on the usability of the various models, both today and in the future. This can be examined from many different perspectives:

- **Database Platform Support:** While section “[Formalism and Theory](#)” considered the theoretical basis for GrC, when one moves from theory to practice in GrC, a seamless transition to the corresponding database management system for a chosen data model would be very useful. In that regard, RDM has the definite advantage, with a wide variety of commercial (Oracle, SQL Server, Informix, etc.) and Open Source (MySQL, PostgreSQL, Ingres, etc.) products available. XML is also a strong player in this regard, since it is a dominant technique for information exchange (data interoperability) among databases, particularly across networked and distributed database solutions. Both RDF and OWL are also demonstrating emergent database capabilities, with the use of SPARQL to query content akin to SQL in RDM.
- **Future Potential:** XML has a potential for future usage in GrC as an information model, due to its widespread and growing usage. In fact, there are many open source XML databases that are starting to be released, including Apache Xindice, Senda XML DBMS, X-Hive/DB, etc. Many of these are in their earliest release stages. For RDM, SQL3, which will include object-oriented extensions, is still being considered from a standard perspective. These object-oriented capabilities are critical to allow granules to be formed in a representation similar to the way that they occur in the “real world,” rather than having the granule be flattened into a relational table. Once approved, the relational database vendors will have to implement SQL3 and release new versions. One can easily hypothesize that XML database systems may make SQL3 obsolete before it even is released. Further, both RDF and OWL repositories, while not specifically database systems with all of their features (concurrency control, transaction processing, recovery, etc.), the ability to query RDF and OWL repositories with SPARQL does make it a viable option for GrC that still falls short of XML.
- **Data Model Translations:** The ABDM as given in section “[The Attribute-Based Data Model \(ABDM\)](#)” has been shown to be capable of

subsuming the features of RDM and FDM (Demurjian and Hsiao 1988) at both a model and system level. The XML format has been utilized by commercial database system products to allow entire databases to be exported as a set of schemas and associated instances, which can be then moved among database platforms and is suitable for use in GrC. While ABDM can be used in place of RDM and FDM models without a loss of information, the modeling capabilities of XML, RDF, and OWL exceed ABDM (RDM and FDM) for data model translation but would suffer as above in regard to being realized with capabilities that rival commercial database products.

Overall, the clear leader from a model perspective is ABDM and its ability to subsume RDM and FDM model capabilities particularly if the underlying database product is factored in. From a practical perspective, while RDM may have the edge in commercial and open source database platforms, using XML (and XMI) to easily translate data across relational database platforms or be used by a custom application that would work with the XML schema and instances. The future seems pointed toward XML for sharing and database usage, particularly if more commercial database products based on XML become available and more widespread in usage. At the present, RDF and OWL are more restricted; while SPARQL mimics SQL to access RDF and OWL repositories, there is none of the needed database functionality (security, recovery, concurrency, etc.).

Future Directions

In this entry, candidate information models and their suitability for granular computing (GrC) have been explored. This entry considered six different models, presented in section “[Candidate Information Models](#)”: the attribute-based data model (ABDM) (Hsiao and Harary 1970), the relational data model (RDM) (Codd 1970), the functional data model (FDM) (Shipman 1981), the extensible markup language (XML) ([World Wide Web Consortium](#)), the resource description

framework (RDF) (Powers 2003), and the web ontology language (OWL) (Allemang and Hendler 2011). Two of these models (RDM and XML) are dominant from a database system perspective today, but all six models have different capabilities to offer when modeling, reasoning, and analyzing for GrC. To place the six models in their proper perspective, in section “[Suitability of Information Models for Granular Computing](#),” the models were explored using three different criteria: Formalism and Theory, Expressive Power, and Extensibility and Utility. For the first criterion (section “[Formalism and Theory](#)”), each of the six models has a strong formal basis and long history of usage; to varying degrees all are suitable for a formal basis for GrC. For the second criterion (section “[Expressive Power](#)”), we compared based on three factors: grouping capability, linguistic representation of granules, and constraint and type checking. In the analysis, ABDM, RDM, FDM, RDF, and OWL are very comparable in what they offer, but XML has the potential to transcend all three, with both its current capabilities and its emerging characteristics and features. Finally, for the third criterion (section “[Extensibility and Utility](#)”), there were different strengths for different models: for ABDM, its strength was its ability to subsume RDM and FDM at the data model level (Demurjian and Hsiao 1988); for RDM, its strength was in the deployed database platforms (commercial and open source) and the future potential of SQL3 with object-oriented capabilities (which will improve granule representation); and for XML, the usage in relational databases for information exchange, coupled with the arrival of XML database systems, may have the greatest potential for the future; RDF and OWL allow access to repositories via SPARQL but lack key database functions. This leaves us with two recommendations, from a theory perspective, while all six data models are appropriate to some degree for GrC, ABDM and RDM have to be strongly considered due to their long history of usage and formal basis, with FDM an alternative when there is a desire to have a functional/logic-based basis to GrC; and, from a practical perspective, if one is to transition from GrC theory to actual usage, RDM

is the choice today, but XML has an excellent chance to be the choice of the future. Usage of RDF and OWL to add semantics is more farther out, but they are the two models that should be carefully monitored for GrC.

Bibliography

- Allemang D, Hendler J (2011) Semantic web for the working ontologist: effective modeling in RDFS and OWL, 2nd edn. Morgan Kaufmann, Amsterdam
- Codd E (1970) A relational model of data for large shared data banks. *Commun ACM* 13(6):377–387
- Demchenko Y et al (2005) Security architecture for open collaborative environment. In: Sloot P et al (eds) *Advances in grid computing – EGC2005*, vol 3470. LNCS. Springer, Heidelberg
- Demurjian S, Hsiao D (1987) The multi-lingual database system. In: *Proceedings of the 3rd international conference on data engineering*. IEEE Computer Society, Washington, DC
- Demurjian S, Hsiao D (1988) Towards a better understanding of data models through the multilingual database system. *IEEE Trans Softw Eng* 14(7):946–958
- Gray P, King P, Kerschberg L (1999) Functional approach to intelligent information systems. *Special Issue J Intell Inf Syst* 12(2–3):107–111
- Hsiao D, Harary F (1970) A formal system for information retrieval from files. *Commun ACM* 13(2), Corrigenda 13(4):67–73
- Lin T (1988) Neighborhood systems and approximation in relational database and knowledge bases. In: *Proceedings of the fourth international symposium on methodologies of intelligent systems*, poster session. IEEE Computer Society, Washington, DC
- Lin T (1989) Chinese wall security policy – an aggressive model. In: *Proceedings of the fifth aerospace computer security application conference*. IEEE Computer Society, Washington, DC
- Lin T (1992) Attribute based data model and poly-instantiation. In: Aiken R (ed) *Education and society – information processing ‘92*, vol 2, proceedings of the IFIP 12th world computer congress. North-Holland, 1992
- Lin T (1997) Granular computing. In: *Announcement of the BISC special interest group on granular computing*
- Lin T (1998a) Granular computing on binary relations I: data mining and neighborhood systems. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*. Physica, Heidelberg, pp 107–121
- Lin T (1998b) Granular computing on binary relations II: rough set representations and belief functions. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*. Physica, Heidelberg, pp 121–140
- Lin T (2005) Granular computing: a problem solving paradigm. In: *Proceedings of 14th international conference*

- on fuzzy systems, 2005. IEEE Computer Society, Washington, DC
- Lin T (2006) A roadmap from rough set theory to granular computing. In: Wang G et al (eds) Proceedings of first international conference on rough sets and knowledge technology – RSKT 2006, vol 4062. LNCS. Springer, Heidelberg, pp 33–41
- Pawlak Z (1991) Rough sets – theoretical aspects of reasoning about data. Kluwer, Heidelberg
- Powers S (2003) Practical RDF, 1st edn. O'Reilly Media, Beijing
- Rothnie JB Jr (1974) Attribute based file organization in a paged memory environment. *Commun ACM* 17(2):63–69
- Shipman D (1981) The functional data model and the data language DAPLEX. *ACM Trans Database Syst* 6(1):140–173
- Skowron A, Stepaniuk J (2001) Information granules: towards foundations of granular computing. *Int J Intell Syst* 16(1):57–86
- Smith J, Smith D (1977) Database abstractions: aggregation and generalization. *ACM Trans Database Syst* 2(2):405–413
- Wang D et al (2004) Medical privacy protection based on granular computing. *Artif Intell Med* 32(2):137–149
- Wong E, Chiang T (1971) Canonical structure in attribute based file organization. *Commun ACM* 14(9):593–597
- World Wide Web Consortium. <http://www.w3c.org/XML/>. Accessed 10 July 2008
- XML 1.0 Fourth Edition Specification. <http://www.w3.org/TR/2006/REC-xml-20060816/>. Accessed 10 July 2008
- Zadeh L (1996) Fuzzy logic = computing with words. *IEEE Trans Fuzzy Syst* 4(2):103–111
- Zadeh L (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 90(2):111–127
- Zadeh L (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Comput* 2(1):23–25



Introduction to Granular Computing

Tsau-Young Lin

Institute of Data Science, GrC Society, San Jose, CA, USA

Article Outline

Granular Computing from Rough Set and Generalized Theory
Granular Computing in Database Theory
Granular Computing in Social Networks
Granular Computing and Fuzzy Set Theory
Grid/Cloud Computing: A New Addition to GrC
General Issues in Granular Computing
Conclusion
Bibliography

In this edition, this editor will not impose his view onto the contributed articles, so we simply list the article in alphabetical order of the title and identify its related papers that have appeared in the previous edition. Here are the articles.

- 1) A Theory of Approximate Computing for Numerical Analysis by Lin, T. (see Scott 2016, Chap. 1, p. Third bullet.)
 - 2) Algebraic Models and Granular Computing by Wasilewska A. (see Item 30.)
 - 3) (fuzzy) Fuzzy Function approximation property of fuzzy systems and its error analysis by Li h.
 - 4) (fuzzy) Fuzzy System Representations of Stochastic systems by Li h (whose the four-stage inverted pendulum is a key event in the history of fuzzy theory.)
 - 5) (fuzzy) Fuzzy Similarity for Parallel Function Computation Model by Chang C. (He coined the term fuzzy topology.)
 - 6) Information System Design Using Fuzzy and Rough Set Theory by Petry F. (see Item 23.)
 - 7) Interval Computing by Kearfott, R. (In 1), we consider the computing of a collection of intervals (which is a local base of new topology on the reals).
 - 8) Knowledge Engineering from Email Archives by Srinivasan, A., Chan, C.
 - 9) Mereology by H. Tsai (This paper reformulates mereology in rough set theory.)
- Security group: Papers 10)–14) and references (Lin and Vachon 2011; Lin 1989c; Lin et al. 1989) are included in this group. Observe that 9 permission bits in UNIX induce very naturally a granulation (generalized partition) on UNIX files; access control naturally induces a granulation on all un files that can be analyzed by granular computing.
- 10) Access Control for XML Big Data Applications by De la Rosa Algarin, A., Demurjian, S., Jackson, E.
 - 11) Issues in Access Control and Privacy for Big Data by Osborn, S., Servos, D., Shermin, M.
 - 12) Privacy Preservation in Big Data Analytics by Tsai, Y., Wang, S., Hong, T.
 - 13) Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment by Hinke, T., Shaw, D.
 - 14) Security with Privacy by Bertino, E.
 - 15) Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm (Miyamoto, S.; rough set theory).
 - 16) Variable Precision Approximations in Rough Sets (Ziarko, W.; rough set theory).
- Paper 17) to paper 39) which were in the first edition have been updated (except Item 20), 21), 24), and 34).
- We will republish the original introduction as a guide to the original group of papers; all of these papers have been updated by original authors (all authors have expressed that the article, including those that have made no actual change, is up to

date, except four articles; one of them was by Professor Zadeh; this editor believes that the idea is still current.)

Here is the original introduction:

What is granular computing (GrC)? It is a shifting paradigm. Let us start with a few words about how the term was coined.

In the academic year 1996–1997, when Lin (this section editor) took his sabbatical leave at UC-Berkeley, Zadeh suggested granular mathematics (GrM) as his research area.

To limit the scope, Lin proposed the term *granular computing* (Zadeh 1998a). What was GrC at that time? Zadeh had outlined it in his 1997 seminal paper (Zadeh 1998b). Lin took an incremental approach:

He mapped his neighborhood system (Lin 1988) to Zadeh’s intuitive definition (Zadeh 1996) and used it as his First GrC model (Lin 1998a, 1998b, 1999). It may be important to point out that the concept of neighborhood systems, which was motivated from approximate retrieval in databases (Lin 1989b), is a generalization of the topological neighborhood system that formalizes the ancient intuition of infinitesimal granules.

Much progress has been achieved since then. This section has been organized to represent this progress and to reflect the current state of GrC.

We believe in the incremental approach, namely, that each new step is based on solid results and moves forward. So many special theories and applications are gathered here. Jointly, they reflect the current state of GrC and may also implicitly hint at the ultimate goals of GrC.

To grasp the main idea from such a diverse collection of papers, a roadmap will be helpful: We suggest the reader start with the first four sections of T.Y. Lin’s paper *Granular Computing: Practices, Theories, and Future Directions*. There, the reader may want to pay special attention to the first three examples:

E1)	Human body is granulated into head, neck, etc.
-----	--

So far, there is no flawless formal model. Obvious models do not work satisfactorily; we need a much more subtle theory.

E2)	Space-time is granulated intuitively into infinitesimal granules.
E3)	The Heisenberg uncertainty principle.

The last two examples have been fully digested by mathematicians and scientists. There are two solutions to the first example, namely, topology and nonstandard analysis, so the readers may want to skim through Lin’s Section “Second GrC Models and Modern Examples” (about neighborhood system/pretopology) in.

- 17) *Granular Computing: Practices, Theories, and Future Directions* and the two articles.
- 18) *Non-standard Analysis, An Invitation to* by Wei-Zhe Yang.
- 19) *Granular Computing and Modeling of the Uncertainty in Quantum Mechanics* by Kow-Lung Chang. These are classical subjects; the authors are a mathematician and a physicist, respectively. Having digested these readings, the reader may proceed to Zadeh’s.
- 20) (fuzzy) *Fuzzy Logic* to gain some feeling about the modern view, and maybe use it to examine the very first example, E1. The article is full of revolutionary ideas; it is, we believe, one of his best papers in presenting an overview of his idea.

Here we paraphrase or quote some of his assertions from his article: “There are many misconceptions about fuzzy logic. A common misconception is that fuzzy logic is fuzzy. In reality, fuzzy logic is not fuzzy. Fuzzy logic deals precisely with imprecision and uncertainty. In fuzzy logic, the objects of deduction are, or are allowed to be, fuzzy, but the rules governing deduction are precise.”

Fuzzy logic is much more than a logical system. More specifically, fuzzy logic has many facets. “The principal facets are: the logical facet, FLI; the fuzzy-set-theoretic facet, FLs, the epistemic facet, Fle; and the relational facet, FLr.”

To gain a glimpse into the nature of fuzzy logic, we examine some examples taken from his article. The left-hand column is “the familiar example of deduction in Aristotelian, bivalent logic. In this example, there is no imprecision

and no uncertainty.” On the other hand, the right-hand column is an example of fuzzy logic which is “in an environment of imprecision and uncertainty.”

$$\begin{array}{l} \text{\$} \text{\textbackslash equal} \{ \text{\&} \text{\textbackslash text} \{ \text{all men are mortal} \} \text{\&} \text{\textbackslash text} \{ \text{most} \\ \text{swedes are tall} \} \text{\textbackslash cr} \text{\&} \text{\textbackslash text} \{ \text{Socrates is a man} \} \text{\&} \text{\textbackslash text} \\ \{ \text{Magnus is a swede} \} \text{\textbackslash cr} \text{\&} \text{\textbackslash overline} \{ \text{\textbackslash text} \{ \text{Socrates is} \\ \text{mortal} \} \} \text{\&} \text{\textbackslash overline} \{ \text{\textbackslash text} \{ \text{it is likely that Magnus is} \\ \text{tall} \} \} \text{\textbackslash cr} \} \text{\$} \end{array}$$

To deduce the answer from the premises (for the right-hand example), “it is necessary to precisiate the meaning of “most” and “tall,” with “likely” interpreted as a fuzzy probability which, as a fuzzy number, is equal to “most.” This simple example points to a basic characteristic of fuzzy logic, namely, in fuzzy logic precisiation of meaning is a prerequisite to deduction.” In this example, “deduction is contingent on precisiation of “most,” “tall,” and “likely.” The issue of precisiation has a position of centrality in fuzzy logic.” “In fuzzy logic, the deduction is viewed as an instance of question-answering.”

[Digression] To this point, we may want to observe that there is some similarity to Lin’s context-based reasoning method, which is a derivative of $\{(\epsilon, \delta)\}$ -definition; see Lin’s article “Second GrC Models and Modern Examples” in *Granular Computing: Practices, Theories, and Future Directions* on the Meaning of “Near.”

To avoid misrepresentation, we will not offer a summary here, but urge the reader to read Zadeh’s article.

There are five areas that have great interactions with GrC, namely (in alphabetical order), databases (especially data mining), fuzzy theory (especially fuzzy control), grid/cloud computing, rough set theory (implicitly extended to topology), and social networks. We will group these articles (plus general issues) accordingly. We shall start from rough set theory.

Granular Computing from Rough Set and Generalized Theory

Briefly, rough set theory (RST) is a theory of equivalence relations (which are equivalent to partitions), so the RST community views GrC as

a generalized partition theory. Recall that an equivalence relation is a reflexive, symmetric, and transitive binary relation. Thus, the “next” generalizations are the tolerance relation (dropping transitive), the partial ordering relation (dropping symmetric), and the fuzzified equivalence relation.

We have collected the corresponding articles:

- 21) *Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems* by Lech Polkowski.
- 22) *Granular Computing and Data Mining for Ordered Data: The Dominance-Based Rough Set Approach* by Salvatore Greco, Benedetto Matarazzo1, and Roman Slowinski.
- 23) *Rough and Rough-Fuzzy Sets in Design of Information Systems* by Theresa Beaubouef and Frederick Petry.
- 24) *Multi-Granular Computing and Quotient Structure* by Ling Zhang and Bo Zhang.

Thanks to Professor Polkowski’s diligent search, we are aware that the term “tolerance relation” first occurred much earlier than one might expect. However, in the rough set community, Nieminen (1988) and Lin (1989b) were probably the first to use the terms “tolerance equality” and “neighborhood of tolerance,” respectively.

In fact, progress on the tolerance relation is quite far reaching; it is a complete Pawlak Theory (CPT); it is implicitly in (Lin 2006).

Recall that an RST is called a CPT, if it includes analogous theories covered in Pawlak’s book (except logic), namely, a theory of approximations and a complete knowledge representation system; the latter notion is defined below (Definition 1).

There are more results in this direction: Let $(\mathcal{U}, \beta)$ be a Global GrC Model with β being a full covering. Then the model $(\mathcal{U}, \beta)$ is a CPT, if β is a semigroup under intersection. However, there are counter examples, namely, some GrC on partial ordering cannot be a CPT.

The quotient structures, addressed by Zhang-Zhang, have been neglected by the rough set community, except some brief results by Lin. Quotient structures are essential for information hiding and knowledge representations.

The quotient structure of a partition is simply a classical set. Zhang-Zhang actually has considered topological cases, and Lin has considered pretopological cases (neighborhood systems).

In general, it can be much more complicated. We shall conclude this paragraph with an example from algebraic geometry:

Let the universe U be the polynomial ring over the complex numbers, and β be the collection of all prime ideals. Then the quotient structure is the complex line plus a generic point; see “Granular and Related Structures” of Lin’s article *Granular Computing: Practices, Theories, and Future Directions* for more details.

Granular Computing in Database Theory

Relational databases can be related to GrC in two views, Fifth or Fourth GrC models (Relational or Multibinary GrC Models).

A Fourth GrC Model $\{GDM_1 = (U, \beta)\}$ is called a Granular Data Model (GDM), if the granular structure $\beta = \{R_1, R_2, \dots\}$ is a collection of equivalence relations. By giving each equivalence relation and its equivalence classes meaningful names, GDM becomes a relation instance, called an information Table (IT) in RST. This naming process is called a knowledge representation. Let $\{IT_1 = (U, A_1, A_2, \dots)\}$ be a knowledge representation of GDM_1 , where A_i , called an attribute, denotes the meaningful names of R_i . Pawlak observed a beautiful phenomenon:

Theorem 1: $\{(U, R_1, R_2, \dots) \rightarrow (U, A_1, A_2, \dots)\}$, up to isomorphism.

From this theorem, we define:

Definition 1: A knowledge representation of the Fourth GrC model is said to be a complete knowledge representation if the representation (the process) induces an equivalence. In other words, the knowledge representation can recapitulate the model.

Based on this theorem, RST can transform data mining on IT to granule-processing, namely, granular computing, on GDM. This view clarifies the nature of automated data mining and simplifies its computations. For example, the association rule

mining is reduced to counting the cardinality of set theoretical expressions of granules. However, we should also observe that in such computations the semantic part of attribute values is ignored. In general, the semantics of attribute values are not utilized in the RST/GrC approach. So some data-mining techniques, such as clustering, cannot be approached by RST, since clustering uses the metric of the ambient space of attribute values. In this respect, GrC may use techniques in computing with words to deal with semantic issues. However, we should also note that computing with words is still in its inception stage.

In this group, we have collected the following articles,

- 25) *Granular Computing, Information Models* for by Steven A. Demurjian, Here, we would like to note that Steven A. Demurjian’s article views GrC from the database, not the RST, perspective.
- 26) *Rule Induction, Missing Attribute Values and Discretization* by Grzymala-Busse.
- 27) *Cooperative Multi-hierarchical Query Answering Systems* by Zbigniew W. Ras, Agnieszka Dardzinska.
- 28) *Dependency and Granularity in Data-Mining* by Tsumoto-Hirano.
- 29) *Rough Set Data Analysis* by Shusaku Tsumoto.
- 30) *Granular Model for Data Mining* by Anita Wasilewska, Ernestina Menasalvas.

Granular Computing in Social Networks

By interpreting a granule as an “ordered set” or a tuple (in some relations), Lin formalizes Social Networks into a Fifth GrC model, called the Relational GrC Model. Articles.

- 31) *Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities* by Steve Webb, James Caverlee, and Calton Pu and.
- 32) *Social Networks and Granular Computing* by Churn-Jung Liao investigate two issues in social networks: security and positional equivalence.

This area actually started much earlier. In the mid-1970s, Atkin started to investigate mathematical structure in human affairs (Atkin 1974). In fact, he reached a simplicial complex in (Atkin 1977), which is symmetric Fifth GrC model or equivalently a Second GrC model (Global GrC model).

Granular Computing and Fuzzy Set Theory

Basically, any assertion about classical sets can be fuzzified, so the intention here is to address GrC applications that extend beyond set theory.

A fuzzy set is defined by a membership function, which is a bounded nonnegative real-valued function. So it is rather natural to investigate the generalization of fuzzy set theory to general function theory; even to more general cases, such as random variables (measurable functions) and to generalized functions (such as Dirac Functions). So, we have proposed a Sixth GrC model, which is a function-based GrC model.

In most known applications, we often need some extra properties on the granular structure, such as the universal approximation property. Using Banach space's language, it implies that the granular structure is a Schauder base. The collection of membership functions used in fuzzy control and the activation functions in neural networks (e.g., Radial-Basis-Functions), all have such a property.

For this group, we have selected four papers:

- 33) (fuzzy) *Genetic-Fuzzy Data Mining Techniques* by Tzung-Pei Hong, Chun-Hao Chen, and Vincent S. Tseng.
- 34) (fuzzy) *Fuzzy System Models Evolution from Fuzzy Rulebases to Fuzzy Functions* by I.B. Turksen.
- 35) *Granular Neural Network* by Yan-Qing Zhang.
- 36) (fuzzy) *Fuzzy Probability Theory* by Michael Beer.

We should note that many soft computing papers can be classified in this category; however, it has its own section, so we recommend that the

readers visit that section. We specifically recommend reading some papers on Type II fuzzy sets, whose membership functions can be viewed as granule-valued (fuzzy-number-valued) membership functions.

We also recommend that readers compare Michael Beer's *Fuzzy Probability Theory* and nonstandard probability theory (in Wei-Zhe Yang's article). From them, the reader might draw his own conclusions as to what a granular probability (granule-valued probability) theory might be. One could ask even further questions, such as what is a granular function (granule-valued function)?

Grid/Cloud Computing: A New Addition to GrC

In the end, we would like to note that a new technology called cloud computing (a subset of grid computing) is intrinsically a technology of granular computing. This is a most natural example of a Seveth GrC model (a Turing machine-based GrC Model), even though the name GrC is foreign to the experts in this field. We are expecting, in the near future, that this area will influence GrC development the most.

General Issues in Granular Computing

Two articles,

- 37) *Granular Computing, Principles and Perspectives of* by Jianchao Han and Nick Cercone and.
- 38) *Granular Computing, Philosophical Foundation for* by Zhengxin Chen, are about general views on GrC. However, they basically address First, Second, and Third models. Professor Chen's three-dimensional views, philosophical, technical, and social/application dimensions, are quite intriguing and may trigger more general investigations into all of GrC. In this group, we will add Kow-Lung Chang, Tsau Young Lin, Wei-Zhe Yang, and Lotfi A. Zadeh's articles.

Conclusion

As GrC is an emerging field, many directions are possible (Zadeh 1979, 1998a, 1998b; Lin 1989a pre-CrC, (Lin 1989b, 1998a, 1998b); Bargiela and Pedrycz 2002; and Jankowski and Skowron 2007). Some important works, due to lack of time and some other factors (including ignorance of this section editor), may not be included here. We apologize for the omissions, with the hope that many of them are referenced by some of the collected articles.

In editing this section, we have adopted an incremental approach: All papers are scientific papers; the only visionary paper is Lotfi Zadeh's "Fuzzy Logic"; and even in that paper, many of the visions are already accomplished.

Looking at the collection, we believe GrC is a promising field; here are some indications:

1. The recent emergence of cloud computing to GrC (our view) may give GrC a momentum to the real-world applications.
2. From the abstract point of view (category theory-based models), GrC and databases have the same abstract structures. This implies that GrC as an abstract concept can be technologically realized.
3. GrC has many links to various branches of mathematics, namely, algebraic topology (simplicial complex), homological algebra (extension functor), and algebraic geometry (Spec(R)). Topological spaces, probability, and belief functions. These facts indicate that, mathematically, GrC contains some important common structures that are touched by many distinct mathematical fields.

Finally, we would like to extend our thanks to all of the authors for their support of this project, to Professor Zadeh for his generous guidance and advice, and to Dr. Robert Meyers for the opportunity to edit this section.

Bibliography

Atkin RH (1974) Mathematical structures in human affairs. In: Heinemann education book. Pitman Press, Bath. ISBN 0 435 820257

- Atkin RH (1977) Combinatorial connectivities in social systems. Birkhäuser, Basel
- Bargiela A, Pedrycz W (2002) Granular computing. Kluwer, Boston
- Jankowski A, Skowron A (2007) Toward rough-granular computing. In: Proceedings of the 11th international conference on rough sets, fuzzy sets, data mining, and granular computing, (RSFDGrC'07). Lecture Notes in AI. Springer, Toronto, pp 1–12
- Lin TY (1988) Neighborhood systems and relational database. In: Proceedings of CSC'88. ACM Press, New York, p 725
- Lin TY (1989a) Chinese wall security policy – an aggressive model. In: Proceedings of the fifth aerospace computer security application conference., December 4–8, 1989. Kluwer 2004, pp 286–293
- Lin TY (1989b) Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings of the fourth international symposium on methodologies of intelligent systems (poster session), pp 75–86
- Lin T (1989c) Chinese wall security policy-an aggressive model. ACSAC, Tucson, pp 282–289
- Lin TY (1998a) Granular computing on binary relations I: data mining and neighborhood systems. In: Skoworn A, Polkowski L (eds) Rough sets in knowledge discovery. Physica-Verlag, Heidelberg, pp 107–121
- Lin TY (1998b) Granular computing on binary relations II: rough set representations and belief functions. In: Skoworn A, Polkowski L (eds) Rough sets in knowledge discovery. Physica-Verlag, Heidelberg, pp 121–140
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh L, Kacprzyk J (eds) Computing with words in information/intelligent systems. Springer Lecture Notes in AI. Physica-Verlag, Heidelberg, pp 183–200
- Lin TY (2006) A roadmap from rough set theory to granular computing. RSKT, pp 33–41
- Lin T, Vachon P (2011) Secure information flow and file movements: A topological theory of discretionary access controls. IEEE BigData 2017: 1821–1829y Press
- Lin T, Kerschberg L, Trueblood R (1989) Security algebras and formal models: using petri net theory. DBSec:75–96
- Scott R (2016) Numerical analysis. Princeton University
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) Advances in fuzzy set theory and applications. North-Holland, Amsterdam, pp 3–18
- Zadeh LA (1996) The key roles of information granulation and fuzzy logic in human reasoning. In: IEEE International Conference on Fuzzy Systems, September 8–11, 1
- Zadeh LA (1998a) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. Soft Comput 2:23–25
- Zadeh LA (1998b) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–127

Granular Computing and Modeling of the Uncertainty in Quantum Mechanics

Kow-Lung Chang
Physics Department, National Taiwan University,
Taipeh, Taiwan

Article Outline

Glossary
Definition of the Subject
Introduction
Quantum Postulates and Associated Propositions
Dirac's Bra and Ket Notations and the Realization of Wave Function
Realization of q -Representation in Quantum Mechanics
Fourier Transformation and p -Representation
Quantum Uncertainty and Non-compatibility of Observables P and X
Conclusion
Future Directions
Bibliography

Glossary

Newtonian mechanics The classical mechanics based on Newton's law of motion.
Uncertainty Also commonly referred to as error; deviation from the average value.
Extrinsic uncertainty The uncertainty due to the systematic error and random error.
Intrinsic uncertainty The uncertainty due to the nature of particle-wave duality in the quantum system.
 $\psi(\mathbf{r})$ q -representation of the quantum state, called state function or wave function in coordinate space.

$\varphi(\mathbf{p})$ p -representation of the quantum state, called the state function or the wave function in momentum space.

$\rho(\mathbf{r})$ Probability density in coordinate space, defined as $\rho(\mathbf{r}) = \psi^*(\mathbf{r})\psi(\mathbf{r}) = |\psi(\mathbf{r})|^2$

$\tilde{\rho}(\mathbf{p})$ Probability density in momentum space, defined as $\tilde{\rho}(\mathbf{p}) = \varphi^*(\mathbf{p})\varphi(\mathbf{p}) = |\varphi(\mathbf{p})|^2$

$|\psi(\mathbf{r})|$ Probability amplitude of state $\psi(\mathbf{r})$.

Particle-wave duality A quantum object exhibits both the nature of a particle and a wave.

Wave vector $\mathbf{k} = \mathbf{p}/\hbar$.

Wave number One-dimensional wave vector, $k = p/\hbar$.

Quantum state The physical state in a sub-atomic quantum system.

0th Postulate of quantum mechanics Regarding quantum states as elements of Hilbert space $\mathcal{H}$.

1st Postulate of quantum mechanics Assigning each dynamical variable in a quantum system a unique linear Hermitean operator in Hilbert space $\mathcal{H}$.

2nd Postulate of quantum mechanics The set of eigenvectors of a given observable forms the bases of a Hilbert space.

3rd Postulate of quantum mechanics Poisson brackets in classical mechanics are replaced by commutators of the corresponding observables according to the relations in Eq. (5)

$[R, S]$ Commutator of operator R and operator S defined as $RS - SR$.

$\{R, S\}$ Anti-commutator of operator R and operator S defined as $RS + SR$.

Eigenvalue and eigenvector For operator A acting upon a particular state ψ_a , such that $A\psi_a = a\psi_a$, then the scalar number a and the state ψ_a are called respectively the eigenvalue and eigenvector of operator A .

Inner product A numerically valued function of the ordered pair of vectors ψ and φ , denoted by (ψ, φ) such that $(\psi, \varphi) = (\varphi, \psi)^*$. In Dirac's notation, it takes the forms $\langle\psi|\varphi\rangle = \langle\varphi|\psi\rangle^*$.

Hilbert space A complete vector space with norm defined as the inner product.

Dual space The space formed by the set of all functionals satisfying the linearity conditions.

Hermitean operator An operator which is self-adjoint; that is, an operator A equals its adjoint conjugate, $A = A^+$.

Adjoint conjugate operator For an operator A and a pair of vectors ψ and ϕ , the adjoint operator to A , denoted by A^+ if it satisfies the relation $(\psi, A\phi) = (\phi, A^+\psi)^* = (A^+\psi, \phi)$.

$|\psi\rangle$ A ket vector in Hilbert space.

$\langle\psi|$ A bra vector in dual space.

Compatible observables Physical observables that commute to each other.

Projection operator $|a_i\rangle\langle a_i|$ a Hermitean operator that projects any vector in $\mathcal{H}$ onto a subspace.

Closure relation Direct sum of all project operators equals identity operator.

δ -function A distribution function denoted by $\delta(x - a)$ such that

$$\begin{aligned} \int_D f(x)\delta(x - a)dx &= f(a) \quad \text{and} \\ \int_D f(x)\delta'(x - a)dx &= -f'(a) \quad \text{if } a \in D, \\ \int_D f(x)\delta(x - a)dx &= 0 \quad \text{if } a \notin D. \end{aligned}$$

Cauchy-Schwarz inequality Given vectors $|\alpha\rangle$ and $|\beta\rangle$ such that $\langle\alpha|\alpha\rangle\langle\beta|\beta\rangle \geq |\langle\alpha|\beta\rangle|^2$.

G_d Deviation operator with respect to observable G defined as $G_d = G - \langle G \rangle I$

atomic level, the laws of Newtonian mechanics cease to apply and instead a new theory, quantum theory, is introduced to account for the quantum phenomena of the subatomic system.

Uncertainty in the measurement of a quantum system comes from two origins. The first is called the extrinsic uncertainty. Quantum measurements, just as in any scientific measurement, are never exact. There exists a systematic error or uncertainty as well as random ones during the processes of measurement for various reasons such as inaccurate calibration of instruments, limitation in resolution of apparatus or meters, variations in temperature or humidity or even the position adopted by the individual observer in reading data, etc. The second is called the intrinsic uncertainty which arises only in the case of a quantum mechanical system.

This particular quantum uncertainty is inherent in the very nature of particle-wave duality [1] for a subatomic quantum system. The intrinsic uncertainty exists always even if the extrinsic uncertainty is eliminated or minimized to the level of negligibility.

In order to describe the particle-wave duality for a subatomic quantum system, one introduces the wave function, or state function, $\psi(\mathbf{r})$ to represent the quantum state. $\psi(\mathbf{r})$, which provides all the dynamic quantities of the system, is related to the probability density by $\rho(\mathbf{r}) = \psi^*(\mathbf{r})\psi(\mathbf{r}) = |\psi(\mathbf{r})|^2$. Therefore, the average position of the quantum object $\langle \mathbf{r} \rangle$ can be obtained through the following integration

$$\langle \mathbf{r} \rangle = \int \rho(\mathbf{r})\mathbf{r}d^3\mathbf{r} = \int \psi^*(\mathbf{r})\mathbf{r}\psi(\mathbf{r})d^3\mathbf{r}. \quad (1)$$

The uncertainty in position, denoted by $\Delta \mathbf{r}$, can then be expressed by

$$\begin{aligned} (\Delta \mathbf{r})^2 &= \int \rho(\mathbf{r})(\mathbf{r} - \langle \mathbf{r} \rangle)^2 d^3\mathbf{r} \\ &= \int \psi^*(\mathbf{r})(\mathbf{r} - \langle \mathbf{r} \rangle)^2 \psi(\mathbf{r}) d^3\mathbf{r}. \end{aligned} \quad (2)$$

Similarly, the uncertainty in momentum, $\Delta \mathbf{p}$ can be written as

Definition of the Subject

For data collection or data acquisition in granular computing or modeling, particular attention should be paid to data that is involved with the position measurement and/or the momentum measurement of a particle in a physical system. For a system that is governed by Newton's law of motion, measurements of the positions and the momenta of a classical mass point can be made as precise as one wishes if the best measurement conditions are available. When we perform physical measurements in microscopic systems at sub-

$$\begin{aligned}
(\Delta p)^2 &= \int \tilde{\rho}(p)(p - \langle p \rangle)^2 d^3p \\
&= \int \varphi^*(p)(p - \langle p \rangle)^2 \varphi(p) d^3p,
\end{aligned} \tag{3}$$

where $\varphi(p)$ is the state function of the quantum system in momentum space.

The quantum uncertainties of the system obey the following relation (Merzbacher 1970):

$$\begin{aligned}
\Delta x \Delta p_x &\geq \hbar/2 \\
\Delta y \Delta p_y &\geq \hbar/2 \\
\Delta z \Delta p_z &\geq \hbar/2,
\end{aligned} \tag{4}$$

where $\hbar = h/2\pi$, and h is called the Planck's constant.

Introduction

Some basic mathematical tools are necessary in order to understand the uncertainty relations of quantum mechanics in a more rigorous sense. As discussed in the last section, it is only a vague answer to the statement that a quantum object possesses both the nature of a point particle and the nature of a wave at the same time, so that one visualizes the quantum object as a matter wave with some sort of distribution in coordinate space as well as in wave vector space. Therefore, neither can one find a localized point quantum object nor a strictly monochromatic wave motion in the quantum system.

Next we shall discuss the relations of the quantum state and the element in Hilbert space as well as the physical observables and the linear Hermitean operator. The connection between the eigenvalue problems and the preparation of a quantum state for a particular measurement of physical observables are also explored in depth. Dirac's notations of the bra and ket vector in Hilbert space are also reviewed and their connection with the q -representation and p -representation of wave functions is derived. The uncertainty relations can therefore formally be proved as the consequence of non-compatibility of two physical observables of position and momentum.

Quantum Postulates and Associated Propositions

0th Postulate of Quantum Mechanics

For every quantum system, there exists an abstract entity called the state (or state function or wave function in q -representation) which provides the information of all the dynamical quantities of the system, such as positions, momenta, energy, angular momentum, spin, and charge. All the possible states $\psi, \phi, \chi, \dots$ etc. of a given quantum system are elements of Hilbert space $\mathcal{H}$.

1st Postulate of Quantum Mechanics

For the measurement of each dynamical variable in the quantum system, such as the total energy of the system or 3rd component of the orbital angular momentum, there associates a unique linear Hermitean operator in Hilbert space $\mathcal{H}$ corresponding to each dynamical variable. The set of linear Hermitean operators in $\mathcal{H}$ are called physical observables, or simply observables.

Dynamical variable	Physical observable
$E(\text{energy})$	$\rightarrow H$ (Hamiltonian operator)
ℓ_3 (3rd component of ℓ)	$\rightarrow L_3$ (3rd component of angular momentum operator L)

The physical quantity a corresponding to the observable of the quantum state ψ is obtained from the inner product of the order pair ψ and $A\psi$, where A is the corresponding linear Hermitean operator associated with dynamical variable a . The inner product $(\psi, A\psi)$ is called the expectation value, or the average value associated with the dynamical operator A for state ψ of the quantum system.

$$a = (\psi, A\psi)$$

In general, the operator A acting on ψ will change ψ into another element ϕ in Hilbert space $\mathcal{H}$, which implies that the action of the measurement usually would disturb the quantum system and brings the original quantum state ψ into a new state ϕ by the external perturbation during the process of measurement, i.e.,

$$A\psi = \varphi.$$

In a particular case, if an operator A , acts on ψ_a such that

$$A\psi_a = a\psi_a,$$

that is, when A acts upon a particular quantum state ψ_a , the resultant state is the same as the one before except multiplying by a scalar number a , then it is said that the quantum system is prepared for the measurement of the dynamical variable associated with the physical observable A . This particularly prepared quantum state ψ_a is called the eigenstate of the operator A . The scalar numerical value a , called the eigenvalue of operator A , is the result of the measurement for the dynamical observable.

Since the value of the physical measurement is always a real quantity, we therefore conclude the first proposition:

Proposition 1 The eigenvalues for a Hermitean operator are real.

For $A = A^+$, and letting ψ_a be the eigenstate of A , then

$$\begin{aligned} a(\psi_a, \psi_a) &= (\psi_a, A\psi_a) = (A^+\psi_a, \psi_a) \\ &= (A\psi_a, \psi_a) = (\psi_a, A\psi_a)^* = a^*(\psi_a, \psi_a) \end{aligned}$$

Since $(\psi_a, \psi_a) \neq 0$, we have $a = a^*$.

It is also ready to show Proposition 2.

Proposition 2 Two eigenvectors of the same Hermitean operator are orthogonal if the corresponding eigenvalues are unequal.

Let

$$A\psi_1 = a_1\psi_1, \quad A\psi_2 = a_2\psi_2.$$

And if $a_1 \neq a_2$, then

$$\begin{aligned} (\psi_1, A\psi_2) &= a_2(\psi_1, \psi_2) \\ &= (A^+\psi_1, \psi_2) = (A\psi_1, \psi_2) = a_1(\psi_1, \psi_2). \end{aligned}$$

Therefore, $(a_1 - a_2)(\psi_1, \psi_2) = 0$, that concludes that ψ_1 is orthogonal to ψ_2 if $a_1 \neq a_2$. Namely $(\psi_1, \psi_2) = 0$.

With these propositions, we can formulate the 2nd postulate of quantum mechanics.

2nd Postulate of Quantum Mechanics

The set of eigenvectors ψ_a for all possible value a of a given Hermitean operator corresponding to a physical observable form the bases of a Hilbert space $\mathcal{H}$.

If there exists a complete set of linearly independent vectors ψ_a which are eigenvectors of both operators R and S , then the two corresponding physical observables, R and S , are said to be compatible, we then have:

Proposition 3 If two observables are compatible, their corresponding operators commute, i.e., if

$$\begin{aligned} R\psi_a &= r_a\psi_a \\ S\psi_a &= s_a\psi_a \end{aligned}$$

then $(RS - SR)\psi_a = [R, S]\psi_a = 0$.

Any vector ψ in $\mathcal{H}$ can be expressed in terms of linear combination of ψ_a , i.e.,

$$\psi = \sum_a \alpha_a \psi_a,$$

one can easily verify that it also satisfies $[R, S]\psi = 0$ for any state ψ in $\mathcal{H}$.

Therefore, the observable R and the observable S commute.

From the point of view of physical measurement, the compatibility of two observables implies that one is able to prepare a single quantum system for the precise measurements of two different dynamical quantities corresponding to R and S respectively at the same time.

Unfortunately, not all physical observables are compatible. The position and the momentum are just an example, and the non-compatibility of position operator X and momentum operator P is in fact the foundation of quantum theory that forms the 3rd postulate of quantum mechanics.

3rd Postulate of Quantum Mechanics

Every Poisson bracket in classical mechanics for canonical variables (p_i, q_i) is replaced by the commutator of the corresponding operators in the quantum system with the following relations

Poisson bracket	Quantum commutator operator
$[q_i, q_j] = 0$	$\rightarrow [X_i, X_j] = 0$
$[p_i, p_j] = 0$	$\rightarrow [P_i, P_j] = 0$
$[p_i, x_i] = \delta_{ij}$	$\rightarrow [P_i, X_j] = \hbar/i\delta_{ij}.$

(5)

The non-commuting of the dynamical observables P_i and X_i will in fact not only lead to the fundamental quantization of the microscopic physical system, it also leads to the deviations in the position measurement and the momentum measurement. The non-commutativity of P and X in the third quantum postulate signifies that one is not able to prepare a quantum state for the simultaneous measurement of momentum and position with absolute precision. Therefore, one shall regard the uncertainty relations of Eq. (4) as the direct consequence of this quantum postulate.

Dirac's Bra and Ket Notations and the Realization of Wave Function

To understand that quantum uncertainty is the direct consequence of two non-commuting physical observables, it is convenient to introduce Dirac's notations of bra and ket vectors. A ket vector, written as $|a\rangle$, represents the abstract quantum state ψ_a , with a inside the notation $|a\rangle$ stands for the eigenvalue of the observable A , and the eigenvalue equation becomes

$$A |a\rangle = a |a\rangle.$$

Similarly we denote the states $\psi, \varphi, \chi, \dots$ etc. respectively by the ket vector notations by $|\psi\rangle, |\varphi\rangle, |\chi\rangle, \dots$ etc.

The bra vectors are the elements in another vector space that are dual to ket vector space.

The bra vector takes the form $\langle\psi|$. The inner product in Dirac's notation is expressed by taking

a pair of dual vectors, i.e., a bra vector $\langle\psi|$ and a ket vector $|\varphi\rangle$, and putting them side by side as $\langle\psi|\varphi\rangle$ of which one always simplifies the notation as $\langle\psi|\varphi\rangle$. The inner product is a linear functional that defined a complex number for every ket vector in Hilbert space $\mathcal{H}$, such that

$$\langle\psi|\varphi\rangle = \langle\psi|(\alpha|\varphi\rangle + \beta|\chi\rangle) = \alpha\langle\psi|\varphi\rangle + \beta\langle\psi|\chi\rangle.$$

The set of bra vectors that define all linear functionals associated with the ket vector in $\mathcal{H}$ forms a dual Hilbert space $\widetilde{\mathcal{H}}$. The inner product is a complex number, i.e.,

$$\langle\psi|\varphi\rangle = \langle\varphi|\psi\rangle^*,$$

because bra vectors and ket vectors are elements of two different Hilbert spaces, and they are dual to each other. The inner product of ψ and φ denoted previously by (ψ, φ) is then replaced by $\langle\psi|\varphi\rangle$. The expectation value of the physical observable A becomes

$$\langle\psi|A|\psi\rangle = (\psi, A\psi).$$

The advantage of using Dirac's notations is in the construction of the projection operator, of which a vector in the vector space is projected onto a subspace. Let us denote the projector operator $\mathcal{P}_a$ by putting the bra vector followed by a ket vector immediately as

$$\mathcal{P}_a = |a\rangle\langle a|,$$

therefore, the vector $|\psi_a\rangle$ is automatically constructed when the projection operator $\mathcal{P}_a$ applies upon any vector $|\psi\rangle$, i.e.,

$$|\psi_a\rangle = \mathcal{P}_a |\psi\rangle = |a\rangle\langle a|\psi\rangle.$$

Let $|i\rangle$ be the set of the eigenkets of observable A , with the discrete eigenvalue a_i , then

$$A |i\rangle = a_i |i\rangle.$$

It is often normalized to the eigenvector such that

$$\langle i|j\rangle = \delta_{ij}.$$

Any vector $|\psi\rangle$ in $\mathcal{H}$ can be expressed in terms of the base vectors with the coefficients $\langle i|\psi\rangle$ in the following linear combination,

$$|\psi\rangle = \sum_i (\langle i|\psi\rangle) |i\rangle = \sum_i |i\rangle \langle i|\psi\rangle.$$

The summation $\sum_i |i\rangle \langle i|$ in fact is just the direct sum of projection operator $\wp_i = |i\rangle \langle i|$, which obeys the following closure relation

$$I = \sum_i \wp_i = \sum_i |i\rangle \langle i|.$$

In many cases, the eigenvalue of observables becomes continuous. Taking observable X , the position operator for instance in the quantum system of interest, and for the sake of simplicity, we shall consider only a one-dimensional case, the eigenvalue equation of this quantum system reads

$$X|x\rangle = x|x\rangle,$$

where $|x\rangle$ is the eigenket of observable X with eigenvalue x which is a continuous parameter representing the position of the quantum object.

The normalization of an eigenvector with continuous eigenvalue is taken as a δ -function

$$\langle x|x'\rangle = \delta(x - x'),$$

and the corresponding closure relation becomes

$$I = \int |x\rangle dx \langle x|.$$

Realization of q -Representation in Quantum Mechanics

Let us introduce a unitary operator

$$U(P; \xi) = e^{-\frac{i}{\hbar} \xi P}$$

where ξ is a real continuous parameter and P is the momentum operator. It is unitary because

$$U^+(P; \xi) = e^{\frac{i}{\hbar} \xi P}$$

and

$$U(P; \xi)U^+(P; \xi) = U^+(P; \xi)U(P; \xi) = I.$$

If one expands $U(P; \xi)$ in terms of a power series in P , i.e.,

$$U(P; \xi) = I + \left(-\frac{i}{\hbar} \xi P\right) + \frac{1}{2!} \left(-\frac{i}{\hbar} \xi P\right)^2 + \dots,$$

and calculates the commutator $[X, P^n]$ term by term, and makes use of

$$[X, P^n] = i\hbar n P^{n-1},$$

which is based on the third postulate of quantum mechanics, then one obtains

$$[X, U(P; \xi)] = \xi U(P; \xi)$$

or

$$XU(P; \xi) = U(P; \xi)(X + \xi I).$$

It can be easily verified that $U(P; \xi)|x\rangle$ is also an eigenket of X with eigenvalue $(x + \xi)$, i.e.,

$$\begin{aligned} XU(P; \xi)|x\rangle &= U(P; \xi)(X + \xi I)|x\rangle \\ &= (x + \xi)U(P; \xi)|x\rangle \quad \text{and} \\ U(P; \xi)|x\rangle &= |x + \xi\rangle, \end{aligned}$$

where we have chosen the phase factor of the new eigenket $|x + \xi\rangle$ to be 1.

Consider the matrix element of the operator, defined as $\langle x|U(P; \xi)|x'\rangle$ for infinitesimal value of ξ , then by keeping up to first order in ξ , one has

$$\langle X|U(P; \xi)|x'\rangle = \langle x|x' + \xi\rangle \doteq \left\langle x \left| \left(I - \frac{i}{\hbar} \xi P \right) \right| x' \right\rangle.$$

With the properties of the δ -function, one concludes

$$\begin{aligned}\langle x|P|x'\rangle &= \frac{\hbar}{i} \lim_{\xi \rightarrow 0} \frac{1}{\xi} \{\delta(x-x') - \delta(x-x'-\xi)\} \\ &= \frac{\hbar}{i} \frac{d}{dx} \delta(x-x') = -\frac{\hbar}{i} \frac{d}{dx'} \delta(x-x').\end{aligned}$$

Therefore, one is ready to establish the connection between the axiomatic quantum postulates and the q -representation in quantum mechanics by defining the wave function $\psi(x)$ to be the inner product of $|\psi\rangle$ and $|x\rangle$, i.e.,

$$\psi(x) = \langle x|\psi\rangle,$$

and q -representation of observable P , denoted by $\wp$ leads to

$$\begin{aligned}\wp\psi(x) &= \langle x|P|\psi\rangle = \int \langle x|P|x'\rangle dx' \langle x'|\psi\rangle \\ &= -\frac{\hbar}{i} \int \frac{d}{dx'} \delta(x-x') dx' \psi(x') = \frac{\hbar}{i} \frac{d}{dx} \psi(x),\end{aligned}$$

namely, the q -representation of the momentum operator becomes the differential operator with respect to coordinate x . The quantum theory formulated in the previous sections can be translated into the q -representation as follows:

Abstract formulation	q - representation
State ψ	Wave function $\psi(x)$
Observable X	Coordinate x
Observable P	Differential operator
	$\wp = \frac{\hbar}{i} \frac{d}{dx}$
Any observable $F(X, P)$	$F(x, \wp) = F\left(x, \frac{\hbar}{i} \frac{d}{dx}\right)$

Fourier Transformation and p -Representation

As we have mentioned, a quantum system can be described either in the coordinate space of x , or in the momentum space (or wave number space) of p . The p -representation of the quantum state $\varphi(p)$ is similarly defined as the inner product of the order pair vectors $|p\rangle$ and $|\psi\rangle$, i.e.,

$$\varphi(p) = \langle p|\psi\rangle.$$

It is of interest to relate $\varphi(p)$ with $\psi(x)$ by inserting an identity projection operator before the state $|\psi\rangle$, i.e.,

$$\varphi(p) = \int \langle p|x'\rangle dx' \langle x'|\psi\rangle = \int \langle p|x'\rangle dx' \psi(x'). \quad (6)$$

Since

$$\begin{aligned}p\langle p|x\rangle &= \langle p|P|x\rangle = \int \langle p|x'\rangle dx' \langle x'|P|x\rangle \\ &= \frac{\hbar}{i} \int \langle p|x'\rangle \frac{d}{dx'} \delta(x-x') = -\frac{\hbar}{i} \frac{d}{dx} \langle p|x\rangle,\end{aligned}$$

one finds, with proper normalization, that

$$\langle p|x\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{-\frac{i}{\hbar}px},$$

and Eq. (6) can be expressed as

$$\varphi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{-\frac{i}{\hbar}px} \psi(x) dx.$$

Conversely, $\psi(x)$ is the inverse Fourier transform of $\varphi(x)$, i.e.,

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{\frac{i}{\hbar}px} \varphi(p) dp.$$

Quantum Uncertainty and Non-compatibility of Observables P and X

According to the third quantum postulate that $[P, X] = \frac{\hbar}{i}I$, where I is an identity operator, one can show that the deviation operators, or the uncertainty operators of P and X , defined by

$$P_d = P - \langle P\rangle I,$$

$$X_d = X - \langle X\rangle I,$$

$$\text{where } \langle P\rangle = \langle \psi|P|\psi\rangle \equiv \bar{p},$$

$$\text{and } \langle X\rangle = \langle \psi|X|\psi\rangle \equiv \bar{x}$$

are the average value of position and momentum respectively and the commutator of these

deviation operators P_d and X_d will also follow the same quantization rule, i.e.,

$$[P_d, X_d] = \frac{\hbar}{i} I.$$

Let us introduce the anti-commutator operator, defined by

$$P_d X_d + X_d P_d = \{P_d, X_d\}, \quad \text{and denote it by } \Sigma,$$

which is a Hermitean operator. Therefore, one can express the product operator of X_d and P_d as

$$\begin{aligned} P_d X_d &= \frac{1}{2} [P_d, X_d] + \frac{1}{2} \Sigma, \\ X_d P_d &= -\frac{1}{2} [P_d, X_d] + \frac{1}{2} \Sigma. \end{aligned}$$

If we define the ket vectors $|\alpha\rangle$ and $|\beta\rangle$ as

$$\begin{aligned} X_d |\psi\rangle &= |\alpha\rangle, \\ P_d |\psi\rangle &= |\beta\rangle, \end{aligned}$$

then the square of position uncertainty $(\Delta x)^2$ as well as the square of the momentum uncertainty $(\Delta p)^2$ can be expressed respectively as

$$\begin{aligned} (\Delta x)^2 &= \langle \alpha | \alpha \rangle = \langle \psi | X_d^2 | \psi \rangle, \\ (\Delta p)^2 &= \langle \beta | \beta \rangle = \langle \psi | P_d^2 | \psi \rangle. \end{aligned}$$

The quantum uncertainty relation takes the expression

$$\begin{aligned} (\Delta x)^2 (\Delta p)^2 &= \langle \alpha | \alpha \rangle \langle \beta | \beta \rangle \\ &\geq \langle \alpha | \beta \rangle \langle \beta | \alpha \rangle = \left\langle \psi \left| \left(-\frac{\hbar}{2i} I + \frac{1}{2} \Sigma \right) \psi \right\rangle \right. \\ &\quad \times \left. \left\langle \psi \left| \left(\frac{\hbar}{2i} I + \frac{1}{2} \Sigma \right) \psi \right\rangle \right\rangle, \\ (\Delta x)^2 (\Delta p)^2 &\geq \frac{\hbar^2}{4} + \frac{1}{4} \left| \left\langle \Sigma \right\rangle \right|^2, \end{aligned}$$

where Cauchy-Schwarz inequality was applied to obtain the right-hand side of the first inequality sign. Since $|\langle \Sigma \rangle|^2$ is also positive definite because

of the Hermiticity property of the operator Σ , we conclude that

$$\Delta x \Delta p \geq \hbar/2,$$

which is the quantum uncertainty in a one-dimensional case. The equality sign in the last equation holds if the conditions $|\beta\rangle = \lambda |\alpha\rangle$ and $\langle \psi | \Sigma | \psi \rangle = 0$ are met. By making use of

$$\begin{aligned} \langle \psi_{\min} | [P, X] | \psi_{\min} \rangle &= \frac{\hbar}{i} \quad \text{and} \quad \left\langle \psi_{\min} \left| \Sigma \right| \psi_{\min} \right\rangle \\ &= 0, \end{aligned}$$

one obtains $\lambda = \frac{i\hbar}{2(\Delta x)^2}$, and the condition for the state function of the minimal uncertainty in a one-dimensional quantum system satisfies the following equations

$$\left(\frac{\hbar}{i} \frac{d}{dx} - \bar{p} \right) \psi_{\min}(x) = \frac{i\hbar}{2(\Delta x)^2} (x - \bar{x}) \psi_{\min}(x),$$

which can be solved with the solution behaving as a traveling Gaussian wave packet as follows:

$$\psi_{\min}(x) = \frac{1}{\sqrt[4]{2\pi(\Delta x)^2}} e^{-\frac{(x-\bar{x})^2}{2(\Delta x)^2} + \frac{i\bar{p}x}{\hbar}}.$$

Conclusion

The uncertainties in quantum mechanics exist not only in the simultaneous measurement of the positions and the momenta as we have treated previously, uncertainties also exist in the simultaneous measurement of any pair of the dynamical variables that are not compatible, or their corresponding physical observables do not commute. For example, one is not able to obtain the data with absolute precision in the simultaneous measurements of any two components of the angular momentum L_i and L_j , because of the non-vanishing commutator $[L_i, L_j] = i\hbar \epsilon_{ijk} L_k$. If we denote the deviation operator for the observable corresponding to angular momentum by

$$L_d = L - \langle L \rangle I,$$

one can easily show that

$$[L_{di}, L_{dj}] = i\hbar \in_{ijk} L_k,$$

and the square of uncertainty in the i th and j th components of angular momentum, denoted by $(\Delta \ell_i)^2$ and $(\Delta \ell_j)^2$, respectively, can be expressed as

$$(\Delta \ell_i)^2 (\Delta \ell_j)^2 = \langle \psi | L_{di}^2 | \psi \rangle \langle \psi | L_{dj}^2 | \psi \rangle.$$

Because of the Hermiticity of L , and making use of the Cauchy-Schwarz inequality, one can derive that

$$(\Delta \ell_i)^2 (\Delta \ell_j)^2 \geq \frac{\hbar^2}{4} \in_{ijk} \bar{L}_k + \frac{1}{4} \bar{K}_{ij}^2,$$

where $\bar{L}_k = \langle \psi | L_k | \psi \rangle$ and $\bar{K}_{ij} = \langle \psi | \{L_i, L_j\} | \psi \rangle$.

The uncertainty in simultaneous measurement of L_i and L_j depends upon the average value of the physical observable of the third component of angular momentum, as well as a positive definite number $\bar{K}_{ij}^2$. Therefore, the minimum of the uncertainty

$$(\Delta \ell_i)(\Delta \ell_j) \text{ becomes } (\Delta \ell_i)(\Delta \ell_j) \Big|_{\min} = \frac{\hbar}{2} |\in_{ijk} \bar{L}_k|,$$

Therefore,

$$\Delta \ell_i \Delta \ell_j \geq \frac{\hbar}{2} |\in_{ijk} \bar{L}_k|,$$

which indicates that the uncertainty relation for the product of two components of a given angular momentum in a quantum system cannot be independent of the average value of the third component. Many peculiar quantum phenomena that would not be happening in classical physics do exist in the subatomic quantum system. Quantum uncertainties corresponding to a pair of non-compatible observables can only be obtained through detailed investigation based on quantum dynamical theory.

Future Directions

The limitations in pursuit of accurate scientific measurement of some data in quantum mechanical systems are the direct consequence of the quantum dynamics in the commutation relations among various physical observables. One is able to minimize but not eliminate the uncertainty in the simultaneous measurement in any pair of dynamical variables of which the commutator of the corresponding quantum operators does not vanish. For instance, a plane wave of one dimension can be prepared in a quantum system to achieve the utmost monochromatic frequency only at the expense of the total accuracy in locality. As we have treated in the previous section, Heisenberg's uncertainty relations in position measurement and momentum measurement can only be minimized by preparing the quantum state of a traveling Gaussian wave packet. In fact we can easily show that the state function in momentum space is also of a Gaussian traveling wave packet due to the property that the Fourier transformation of a Gaussian distribution is again a Gaussian distribution. Therefore, it is imperative that one should execute uncertainty control or error management before performing the measurements for physical observables of non-compatibility in the quantum mechanical system. As we have already demonstrated in the simultaneous measurement of any two components of an angular momentum, the uncertainty can be minimized by preparing a quantum state with the least mean value of its third component.

Bibliography

- de Broglie L (1923) *Nature* 112:540. (1924) Thesis, Paris.
- (1925) *Ann. Phys* 3:22
- Dirac PAM (1958) *The principles of quantum mechanics*, 4th edn. Clarendon Press, Oxford
- Jordan TF (1982) *Linear operators for quantum mechanics*, 2nd edn. Wiley, New York
- Merzbacher E (1970) *Quantum mechanics*, 2nd edn. Wiley, New York



Philosophical Foundation for Granular Computing

Zhengxin Chen

Department of Computer Science, University of Nebraska at Omaha, Omaha, NE, USA

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

The Road to Granular Computing

The Nature of Granule and Granule Computing

Granule Measurement

Granular Structure

Granulation Provides a Unified View for

Intelligent Problem Solving

Relationship with Soft Computing and Natural Computing

Relationship with Fundamental Issues of Computing and Complex Systems Problem Solving

Temporal Aspects of Information Granules

Other Related Developments

Summary

Future Directions

Bibliography

Glossary

Granule As the fundamental concept in granular computing, a granule is a clump of elements drawn together by various criteria such as indistinguishability, equivalence, similarity, proximity, or functionality.

Granulation Granulation refers to the process of forming granules.

Granular computing (GrC) In a broad sense, granular computing is the general term referring to any computing theory/technology which

involves elements and granules, with granule, granulated view, granularity, and hierarchy as its key concepts.

Granular structure Granular structure is the collection of granules, in which the internal structure of each granule is visible.

Granularity The granularity of a level refers to the collective properties of granules in a level with respect to their sizes.

Hierarchy In granular computing, hierarchy captures the ordering of levels.

Fuzzy set and fuzzy logic Unlike a conventional set, in a fuzzy set, a fuzzy membership function is used to define the degree of an element belonging to the set. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth as defined by membership functions. Fuzzy logic contributes to the machinery of granular computing.

Neighborhood system A neighborhood system of a point (an element) in the universe is the nonempty family of subsets (referred to as the neighborhood of that point) associated to it.

Rough set Rough set is a formal approximation of a conventional set, using a pair of sets as the lower and the upper approximations of the original set. Rough sets provide a single-layered granulation structure of the universe.

Definition of the Subject and Its Importance

In the last two decades, granular computing (GrC) has emerged as a major research field for further abstraction, generalization, and unification of granule-based key concepts traditionally scattered throughout a wide range of scientific disciplines, offering new opportunities for systematic studies of challenging computational issues cross multiple disciplines. According to Pedrycz, computing becomes innovative and intellectually proactive endeavor that manifests in several fundamental ways; one of them is that it identifies the essential commonalities between the surprisingly

diversified problems and technologies used there, which could be cast into a unified framework known as a granular world. With the emergence of the unified framework of granular processing, we get a better grasp as to the role of interaction between various formalisms and visualize a way in which they communicate. GrC looks at the aspects of compatibility of such information granules and ensuing communication mechanisms of the granular worlds. Granular computing forms a unified conceptual and computing platform. It is noted that (Yao 2007) granular computing may be viewed as an interdisciplinary study of computations in nature, society, and science and abstract models of such computations, with an underlying notion of multiple levels of granularity. We may extract and study principles, strategies, and heuristics common to all types of problem solving and information processing.

An examination of foundations of granular computing, particularly on its philosophical dimension, is extremely important, because it should reveal the hidden nature of granular computing, shed useful light for future directions, and provide guidelines for researchers working in this area. In order to make this article useful to research community in granular computing, instead of using philosophical/epistemological jargons, we will stay with basic terminology used in granular computing.

An important advantage of studying philosophical foundation of granular computing is to identify what is still missing. In particular, we point out that the future of granular computing is inevitably tied to important features of complex adaptive systems, such as self-organization, learning, evolution, and adaptation, as well as key issues in contemporary artificial intelligence (AI), such as embodiment, emergence, and emotion.

The objective of this article is to provide a fair review on certain important aspects related to philosophical foundation of granular computing. In order to achieve this, a thorough examination is needed: focus not just on the nature of GrC experts in this field have expressed but also what they have actually done and not just about the practices within GrC but also recent developments of related fields. We summarize and analyze existing literature to find out what has led to GrC, what have been

explicitly stated, what are implied, and what are still lacking (or overlooked). Due to the huge amount of literature in this area and diverse viewpoints of researchers, a comprehensive coverage of all perspectives and approaches is impossible. Only selected materials of influential work are included. Materials presented in this article (including the author's own opinions, some of them have never been published so far) are not necessarily an indication that they have been accepted by the GrC community in general or approved by any specific individual.

The recent advent of big data not only presents huge new challenges to research and practice in information technology but also offers tremendous opportunities for GrC to solve problems in uncharted terrains. The coming of big data also necessitates the need for in-depth study of some previously unknown or overlooked features for better understanding on the nature of granules, such as dynamicity, as captured in a newly proposed notion, *dynamic granule*.

Introduction

In order to answer the question of what is meant by "foundations of granular computing," we need to answer the question: What constitute the foundations of granular computing? Just like in the examination of foundations of data mining (where granular computing can play an important role in discovery of hidden knowledge buried in huge amount of data) (Chen 2002), foundations of granular computing can be examined from philosophical, technical, and social/application dimensions.

In general, philosophy of science refers to the study of philosophical assumptions, foundations, and implications of science. Topics to be studied include the character and the development of concepts and terms, propositions and hypotheses, and arguments and conclusions, as they function in science, and the manner and types of reasoning; the formulation, scope, and limits of scientific methods; the implication of scientific methods and models; etc. Philosophy of science is closely related to epistemology; as a field on theory of knowledge, it studies the nature, methods, limitations, and validity

of knowledge and belief (Burkhardt and Dufour 1991).

In order to study the philosophy of science, either study it as a whole or its particular fields, we have to analyze the contents of theories as well as practices of scientists working in these fields, to figure out what important philosophical thoughts are implied. Therefore, study of philosophy of science is not confined by “professional” philosophers; researchers in various science fields can (and should) take care of philosophical aspects of their research discipline in their own hands. This is our attitude taken toward the study of philosophical foundation of granular computing.

All the issues related to philosophical foundation of granular computing boil down to one question: *What is granular computing?* (Or what is the *nature* of granular computing?) A simple question, with complex, possibly conflicting, yet incomplete answers! The philosophical foundation of granular computing requires a holistic examination on the foundations of granular computing as a whole.

Before we present any attempted answers to the magic question of what granular computing is, we have to address the issue of *how to* study it. Note that this is a multifaceted question. First, the general methodology, just like the study of philosophy of science in general, we have to take a retrospective/reflective approach to study what researchers have done in published literature related to granular computing: what is explicitly stated (i.e., what researchers in granular computing states what it is), what is implied (i.e., what researchers assume or think what granular computing might be), as well as what is yet to be addressed.

The next aspect of dealing with the “how to” question is to identify specific issues need to be investigated. The answers to the question of “what granular computing is” can be studied from several aspects, in light of interests from the perspective of philosophy of science but stated in terms of granular computing:

- Character and development of concepts and terms, propositions, and hypotheses:

- Where is granular computing from? How does this new paradigm emerge (or converge from multiple predecessors)? In order to understand what granular computing is, we need a historical examination on the roots of granular computing.
- Now that granular computing has established itself as a research discipline, we need to study its basic terminology such as granule, granularity, granulation, etc. What does each of them mean? Without an understanding of these terms and what behind them, we will not be able to understand the nature of granular computing.
- Manner and types of reasoning and the formulation, scope, and limits of scientific methods:
 - What are the major research issues concerned by granular computing community? These issues form a set of descriptive features which collectively characterize what granular computing is.
 - What kinds of problems in which granular computing can solve and concepts such as granularity and granulation are useful? Universality and diversity of these problems not only demonstrate the power of granular computing but also reveal the nature of granular computing from yet another perspective.
- Implication of scientific methods and models:
 - In order to understand the nature of granular computing, it is equally important to examine how it distinguishes itself from other related disciplines. This leads us to examine its relationship with research fields such as soft computing and cognitive informatics.
 - Finally, it is time to examine a more fundamental issue: how is granular computing related to basic issues of computational modeling and complex problem solving? These more advanced issues help in identifying where granular computing stands in a larger picture and help our understanding of granular computing.

Answers to the questions listed above form the bulk of the rest of this article.

Note that the study of philosophical foundation of granular computing is inevitably intertwined with other dimensions of the foundations of granular computing. For example, one topic of study in philosophy of science involves the implications of scientific methods and models, along with the technology that arises from scientific knowledge, for the larger society (Boyd et al. 1991). This justifies the connection of philosophical foundation with other two dimensions of foundations. In case of granular computing, the study of philosophical foundation requires a close examination on the models and algorithms developed (the task of technical dimension) as well as the implication of various applications (the task of social/application dimension).

The Road to Granular Computing

Granule, granularity, and granulation are basic concepts involved in all aspects of human intelligence. First, it is in the language of our daily life, such as “extra fine granulated sugar.” More profoundly, these concepts also form the foundation of science and technology, including biology, mathematics, etc. For example, mathematics, as the study of the measurement, properties, and relationships of quantities using numbers and symbols, has been dealing with computation problems on such granules for thousand years.

For more recent history, we may consider computerized information systems, such as database management systems (DBMS) or information retrieval systems (IR). For example, an examination of relational database from an information theoretic perspective (Lee 1987) has been considered as an important piece of work related to prehistory of granular computing. In a relational database, tuples form the basic granularity level for information storage and retrieval, yet from the transaction processing perspective, it is the entire relation, rather than its constituent tuples, that serves as the basic granule for commit. In fact, the theory and practice of transaction processing are largely around concepts of granularity and granulation: In traditional DBMSs, transactions are the granules to commit and abort, and in

order to assure this, ACID (atomicity, consistency, isolation, and duration) properties are required. In more advanced DBMS systems such as object-oriented DBMS, due to the significantly increased complexity of transactions, sub-transactions may be used as granules for commit, and accordingly, some of the rigid ACID properties may be relaxed somehow. As a main issue of transaction processing, concurrency control is another showcase of granularity-driven activities: although many concurrency control protocols use individual data items as the granules to perform synchronization, multiple granularity has also been used, where different locking modes are imposed on data at different granularity levels of hierarchy, such as the entire database, a partition of the database, a file, a record, and so on. Today’s big data (featured not just by volume but also by velocity, variety, and veracity (Dong and Srivastava 2013)) requires new ways of thinking on how to look at the various forms of granules faced by data management and analytics. The newly proposed BASE (which stands for basically available, soft state, and eventual consistency) properties for transaction processing of big data (Chen 2017; Dong and Srivastava 2013) in parallel distributed environment signals a fundamental shift of data management in this regard.

Note that what is really interested in granular computing community is what is now usually referred to as *information granule* (Skowron and Stepaniuk 2001), a concept emphasizing its role in *intelligent* information problem solving. Another important aspect that distinguishes granular computing with other areas in information technology is that it goes beyond concepts related to granularity and directly addresses *computational aspects* related to granulations.

As noted in Bargiela and Pedrycz (2002, 2006), human-centered information processing has been pioneered by Zadeh through his introduction of the concept of fuzzy sets in the mid-1960s. GrC arose as a synthesis of insights into human-centered information processing by Zadeh in the later 1990s, and the term was coined by T. Y. Lin in 1997. Below we take a closer look on contributions of several research fields which lead to granular computing.

According to Zadeh (1998), granulation refers to partitioning (crisp or fuzzy) of an object into a collection of granules, with a granule being a clump of elements drawn together by various criteria such as indistinguishability, equivalence, similarity, proximity, or functionality. For example, in rough set theory (Pawlak 1991), the partitioning is based on the criterion of equivalence, while in granular computing, partitioning is based on similarity proximity or functionality (note that a rough set is a formal approximation of a conventional set, using a pair of sets as the lower and the upper approximations of the original set).

A granular variable is a variable which takes granular values. Granulation of a variable (or granular variable), such as age, goes from continuous to quantized and, finally, becomes granulated. Several principal types of granules can be distinguished. It has been noted that in fuzzy logic, granulation is achieved through graduation – “In fuzzy logic everything is or is allowed to be graduated, that is, be a matter of degree or, equivalently, fuzzy” (Zadeh 2007a). In addition, granulation of a function is done by summarization, as shown in the rules such as “If X is large then Y is small” (Zadeh 2007b). On the other hand, developments in rough set theory set another example of the universality of the basic concepts of granule and granulation.

As pointed out in historic remarks (Lin 2006a), while fuzzy set theory and Dempster-Shafer theory are considered as examples of non-partition theories (where the universe is not partitioned), rough set theory and algebraic theory of relational database in the 1980s are examples of partition theory (where the universe is partitioned into subspaces). Rough set theory is about equivalence relation which is a subset of binary relation studied earlier. Rough set theory is about partition which is a subset of covering. Later the concept of neighborhood system (NS) was proposed, which is the generalization of rough set concept in that it incorporates various concepts of topological space, covering, binary relation, as well as α -cuts from fuzzy sets into it. Furthermore, each neighborhood can be regarded as a granule, and because of this, early granular computing research

has neighborhood system as its mathematical model. Therefore, the key in a road map from rough set to granular computing lies in the mathematic concepts of coverings, binary relations, and neighborhood systems, where a covering is a collection of granules (subsets) whose union is the whole space. Neighborhood systems have been used for the study of relational databases, and the granulation structures induced by neighborhood systems have been studied. A binary relation is equivalent to a special type of neighborhood system.

The Nature of Granule and Granule Computing

The Nature of Granule

The concept of granule has been studied in AI community for decades (He 2002). According to Zadeh (2007a, b), a granule is a clump of elements drawn together by various criteria such as indistinguishability, equivalence, similarity, proximity, or functionality. In addition, attributes of a granule include probability measure, possibility measure, verity measure, length, volume, etc. There are numerous assumptions implied from various researchers on the basic notions in granular computing and the concept of granule in particular. In order to have a better understanding about the nature of granules, below we examine some of the hidden assumptions and provide a brief comparison.

- *Granule vs. word:* Traditionally, artificial intelligence (AI) has been considered as a research field of reasoning with individual symbols (a symbol is a token of meaning). In these days, many researchers in computational intelligence have put much emphasis on computing with words (CW). As Zadeh noted, granular computing serves as a basis for the methodology of computing with words (CW) (Zadeh 2007b). Although efforts have been made in this aspect, few researchers in granular computing have focused on issues related to linguistics. In current study of granular computing, semantics is demonstrated not through individual

granules but rather on the grouping of granules. Bargiela and Pedrycz (2002) proposed an information processing “pyramid” consisting of three layers organized by the size of information granules, with the lowest one concerning with numeric processing, the intermediate one concerning large information granules, and highest one devoted to symbol-based processing. But this is just one perspective.

- *Granule vs. attribute or entity (in database modeling)*: In entity-relationship approach, although we talk about attributes of information granules, such as size, capacity, and dimension, in general, granules are not distinguished from each other mainly based on descriptive features (or attributes). Granules have low individuality: as a generalization of equivalence class, a granule is an abstract concept, while an entity, as a first-order citizen in data modeling, is a more concrete thing. Although entities with same type can form an entity set, entity sets are not mainly used to form hierarchical levels (rather, they are the “end” products of conceptual modeling), but granulation is. GrC considers both entities and granules at the same time. On the other hand, an attribute (of an entity) is equivalent to a partition of entities.
- *Granule vs. object in object-oriented (OO) paradigm*: In OO, objects are related together through superclass/class/subclass relationships, and properties can be inherited or overridden. The OO paradigm has an anthropomorphic nature. On the other hand, granule is a general concept directly tied to computational aspects; although granules can also be related together through hierarchies, the relationships between various levels in the hierarchy are much more complex (see later discussion). In addition, although hierarchies play a key role in granular computing, other forms of relationships exist between granules (such as volume of a cylinder and the surface of tat cylinder).
- *Granule vs. cluster (in data analysis/data mining)*: Similarity in cluster analysis is often implied by the ~~topology~~ metric of the ambient space, while granule is often granulated by pre-topology (e.g., neighborhood systems).
- *Granule vs. system vs. graph*: Granularity is a measure of the size of the components, or descriptions of components, that make up a system. A system, either natural or man-made, is a set of interacting or interdependent components forming an integrated whole. A system can also consist of a number of sub-systems. Ideas from systems theory have grown with diversified areas, exemplified by the work of B. Banathy, as well as others. Systems described in terms of large components (i.e., larger granularity) are referred to as coarse-grained; otherwise it is termed as fine-grained. Since the graph data structure is a useful tool to describe systems, it can also be a useful modeling tool for the study of granules (see Lin (2008) for the modeling examples). However, since systems theory is devoted to the study of interrelationship of its components to optimize system performance, it does not address the majority of issues concerned in granular computing.

The Nature of Granular Computing

According to Zadeh (2007a, b), the rationale of granulation could be either imperative (forced) or intentional (deliberate), when the precision is costly and there is a tolerance (for imprecision). We can distinguish different forms of granulation, such as spatial (including interval), temporal granulation, etc. Granular computing is an emerging conceptual and computing paradigm of information processing. It has been motivated by the urgent need for intelligent processing of empirical data that is now commonly available in vast quantities, into a humanly manageable abstract knowledge. In this sense, granular computing offers a landmark change from the current machine-centric to human-centric approach to information and knowledge (Bargiela and Pedrycz 2002, 2006, 2007). According to Lin (2006b), any computing theory/technology that involves elements and granules (subsets or generalized subsets) may be called granular computing. Intuitively, elements are the data, and granules are the basic knowledge. This simple definition has the advantage of unifying research activities done by researchers from diverse background. In addition, this simple definition

allows for exploiting the interesting idea of *granulate and conquer*: a softer version of classical divide and conquer; fuzzy granulate and conquer are the cornerstone of the success of fuzzy controls. A very common technique used in the classical “non-partitioning” recursive call is dynamic programming (Granular computing information center).

According to Bargiela and Pedrycz (2002), the fundamental features of GrC consist of the following: allow for multiple abstraction levels (granularity levels), allow for several methods of traversing various levels of hierarchy (referred to as encoding-decoding mechanisms), as well as allow for nonhomogeneous methods. In addition, Zadeh (2007b) noted that granulation is the core concept surrounded by natural language computation, rough set theory, computational theory of perceptions, as well as granular computing. Therefore, there is a simple formula of “granular computing = ballpark computing.”

As an example of recent development in regard to basic concepts in GrC, the study in Dick et al. (2007) considered two granular worlds which are defined on the same universe of discourse but employ different granulations of this universe. In order to translate information from one granular world to the other, there is a need to regranulate the information so that it matches the information granularity of the target world.

Granule Measurement

As indicated earlier, according to Zadeh (2007a, b), a granule is a clump of elements drawn together by various criteria such as indistinguishability, equivalence, similarity, proximity, or functionality. In addition, attributes of a granule include probability measure, possibility measure, verity measure, length, volume, etc. Therefore, granules can be considered as knowledge, comparing with constituting elements which represent data (Lin 1998b). In addition, granules may be interpreted from three semantic views: a granule is a unit of basic knowledge; a granule is a unit of lacking precise knowledge; and a granule is a subproblem of computing (Pedrycz and Chen 2016). We will get back to the last perspective in a later section.

Here we examine the first two perspectives: granules are knowledge units yet consist of uncertainty. Together, these two perspectives imply an intrinsic property associated with granules: the measurement, which refers to the estimation of the magnitude of some attribute of an object, such as its length or weight, relative to a unit of measurement. Consequently, granularity is a measure of the size of the components. In the simplest form, measures can be made through the size of a granule, which indicates the degree of abstraction, concreteness, or detail, of the granule. In the set-theoretic setting, the size of a granule can be the cardinality of the granule. Various uncertainty theories related to granular computing, such as fuzzy set theory or rough set theory, all have its specific way to deal with measurement.

The need for measurement in studying granules implies the importance of dealing with topology, a key concept underlying many subfields in mathematics, such as mathematical analysis. Take a look, for example, the well-known *Heine-Borel theorem*: Each open covering of a closed and bounded set of real numbers has a *finite* subcovering. The granules involved here include the closed and bounded set, the elements in the original open covering, the elements in the resulting subcovering, etc. Therefore, from a granular computing perspective, we may say that the Heine-Borel theorem is concerned with *reduction* from infinitive to finite or *mapping* from infinitive number of granules to finite number of granules.

There are various kinds of granules, and they can be related together through hierarchy or other relationships or aspects such as measurement. For example, a granule could be a three-dimensional (3D) object or could be the surface of this object. Yet the calculation of the volume of a 3D object can be converted to the calculation of the area of the surface, through various theorems developed in mathematical analysis. Here the computation problem (i.e., the volume) on one granule (i.e., the 3D object) is *converted* to another computation problem (i.e., the area) on another granule (i.e., the surface of the 3D object). This is exemplified in *Green's theorem*, a special case of the more general Stokes' theorem.

As a more recent example concerning the importance of measures for granules, let us consider the case of relational database design theory, as well as its contemporary extension to XML file design. Relations (or flat tables) are the most visible granules in relational database design. But what should consist in a relation? This issue has been addressed by normalization theory, where important normal forms such as third normal form (3NF) or Boyce-Codd normal form have been developed, based on the notion of functional dependency and related concepts. But how could this kind of result be generalized to the case of semistructured XML files? Recent study by Libkin and his research group (Libkin 2007) addressed this issue by proposing to make use of entropy-based measures to determine the quality of the granulation for relational database design and further apply this principle to the XML file design. First the concept of relative information content (RIC) was proposed; subsequently other related measures were proposed, such as the Price. Furthermore, $\text{Price (3NF)} = \frac{1}{2}$. Therefore, the conformation of a “good” granule (such as a relational table) is justified by a quantified approach. Although this and related studies are conducted by researchers outside of GrC community, these results can shed new insight for studying features of granules which have not been considered within the scope of GrC.

Granular Structure

Quotient View: Information Hiding

As noted by Pedrycz and Chen (2016), granular structure is the collection of granules, in which the internal structure of each granule is visible. In this sense, granular structure is a collection of white box granules. In contrast, a quotient structure, as derived from the mathematical concept of quotient space, refers to the collection of granules, in which each granule is regarded as an element (or point) of a set, and the “intersections” among granules are abstract to interactions among points (or elements). Therefore, a quotient structure is a collection of black box granules. This perspective provides a quotient view on the hierarchical

structure consisting of granules at the different levels of details. With details encapsulated at lower levels while only abstract information is shown on higher levels, quotient view-based granular structure supports information hiding.

Earlier in Introduction section, we mentioned that DBMS is a major source behind GrC. In fact, database design is a perfect example of information hiding: the details of the instances are hidden at the more abstract level of schema. The success of the well-known entity-relationship approach lies on the information hiding by focusing on schemas (while not completely throwing away the instances, such as the notion of mapping cardinality involved in relationship sets). More recent development to management of semistructured data necessitates the need for XML schema discovery (see, e.g., Chidlovskii 2001 and He 2002), which in essence is to identify the quotient structure and support information hiding.

The theory of hierarchy provides a multilayered framework based on levels. Mathematically, a hierarchy may be viewed as a partially ordered set. There are two types of partial ordering that can be distinguished, i.e., strong vs. weak dependencies: a neighborhood system X is weakly dependent on Y , if every neighborhood of X is a subset of Y , or strongly dependent, if every neighborhood of Y is a union of the neighborhood of X (Lin 1998a, b; Lin and Zadeh 2004); note that strong = weak in the case of partitions. (More discussion on hierarchy will be given later in the section on emergence.)

A granule in a lower level may be a more detailed description of a granule in a higher level with added information. In the other direction, a granule in a higher level is a coarse-grained description of a granule in a lower level by omitting irrelevant details. Note that although we are talking about granules of various granularities connected together in a hierarchical manner, in general, all granules could be related to each other in a complex network. In this sense, what we call hierarchy in GrC is actually a kind of “information hiding” with detailed network information filtered out.

A taxonomy of types of granularity has been discussed (Jankowski and Skowron 2007). First,

one can identify the main differences in types of granularity based on:

1. Whether the scale matters: arbitrary scale vs. non-scale-dependent granularity
2. How levels (and its contents) in a perspective relate to each other
3. How are perception and (mathematical) representation related, such as based on set theory or mereology

The philosophical implication of this taxonomy has also been examined; for example, it is noted that the difference between scale and non-scale dependency and their formal representations roughly fits with Sowa's epistemic and intentional granularity (Maciel et al. 2016) based on Peirce's three categories.

Granular Computing and Ontologies

The concept of hierarchy plays a key role in many other contexts studied in the last decades, including frame systems in artificial intelligence and object-oriented approaches, just name a few. Here we take a look at the recent surge of research related to ontologies, an issue originated in philosophy (Burkhardt and Dufour 1991) and has received ever-increasing attention in recent years from a much wider community.

Even hierarchies play a significant role in both granular computing and ontologies, the relationship between these two research fields is not well-studied, and this is likely due to the fact that although granular computing sets emphasis on both computation and semantics, the study of ontologies is largely "semantics without computation." Nevertheless, ontological aspects have been occasionally studied by granular computer researchers, including the work on taxonomy of types of granularity summarized in an earlier section (Keet 2006). In addition, granular computing can be used to aid the study of ontologies. One example is depicted in Qiu et al. (2006) where the concepts of domain granule and domain granule lattice were introduced and an algorithm of generating the lattice was proposed to capture the ontology. The relationship between granular computing and ontology can take the opposite

direction as well; for example, Zhou and Liang (2006) described a granular computing model based on ontology.

In addition, Jankowski and Skowron (2007) presented a rough-granular approach which incorporates domain knowledge. This additional knowledge, represented by ontology of concepts, is used to assist searching for features (condition attributes) relevant for the approximation of concepts on different levels of the concept hierarchy defined by a given ontology. Yet, the question of how ontologies of concepts can be discovered from sensory data remains as one of the greatest challenges for many interdisciplinary projects on learning of concepts.

Information Hiding and Emergence

As a basic task of granular computing, granularity conversion is to change views with respect to different levels of granularity (with changing details). However, it would be naïve to believe that relationships between different levels of granulations can be described by mappings.

Information hiding as demonstrated in granulation hierarchy may shed important light in revealing the nature of emergence (Holland 1999), where complex phenomena are produced by interaction of simple, tiny things. Think about the operations of zoom in and zoom out: just like using zoom lenses on a camera, we can show a smaller area of an image at a higher resolution or magnification or a larger area at a lower resolution or magnification. Note that there are apparently two kinds of "zoom in": for example, when we are examining an online interactive map, we may click a particular country on the initial world map, then click a particular region or state, then click a particular city, and so on, to take a close look at particular streets or buildings we are interested. On the other hand, sometimes the so-called zoom in only gives us larger prints of the original map without any additional information added: although this latter type of "zoom in" is useful to senior citizens with vision problems, it is useless for most people. Similarly, we may have two types of zoom out: we can get the entire picture as a whole rather than focusing on a particular part of the picture (with all the details kept), or we

summarize or abstract information not available at the more detailed level.

Apparently, zoom in and zoom out are directly concerned with getting information at different granulation levels. When we “zoom out” from the lower level, we may not necessarily just lose the details at the lower level but also gain insight not demonstrated at the lower level. As a typical example, think about the painting “Gala Contemplating the Mediterranean Sea which at Twenty Meters becomes a Portrait of Abraham Lincoln” by Spanish surrealist Salvador Dali. As the title suggests, at the detailed level is Gala’s picture and the surroundings, while at a coarse granulation level, Lincoln’s portrait emerges. This is where the excitement of emergence comes from, and granular computing can play a critical role in finding hidden rules which produce the emergent phenomena.

Due to the complexity of the networks in which granules are interrelated to each other, it would be questionable to assume that an algorithmic study on the procedures for constructing granules and related computation is a universal and realistic approach in GrC. Just like the more general study of complex adaptive systems, the tasks faced in GrC community are closely related to the task of understanding the nature of emergence. As the movement from low-level simple rules among granules to higher-level sophistication, emergence is an ubiquitous feature of the world around us. The hallmark of emergence is the sense of much coming from little. Complex systems present behaviors by drawing upon populations of relatively “unintelligent” individual agents, rather than single, “intelligent” agent. Yet, as John Holland noted, regardless how mysterious it might appear, emergence can be studied as a scientific discipline, because low-level rules and laws can be discovered, and the movement from low-level rules to higher-level sophistication is what we call emergence. This view suggests a hierarchy of levels from which emergence takes place, and the phenomena of emergence are typically bottom-up in nature. Holland underlines a number of features of emergent (or complex) systems and suggests that emergence is a product of coupled, context-dependent interactions. These

interactions, along with the resulting system, are nonlinear, where the overall behavior of the system cannot be obtained by summing up the behavior of its constituent parts. Persistent patterns at one level of observation can become building blocks for persistent patterns at still more complex levels. At each level of observation, the persistent combinations of the previous level constrain what emerges at the next level. As Holland put it, “This kind of interlocking hierarchy is one of the central features of the scientific endeavor. It will lead us into, and out of, the thorny thicket known as reduction – roughly, the idea that we can reduce explanations to the interactions of simple parts.” Yet, no systematic ways are presented on how to discover much needed laws or rules so that emergence can take place.

Think about one of the most mysterious things we can ever have, namely, life. Regardless whether we are talking about how the first life appeared on the Earth millions or billions of years ago, or how a particular biological species (such as rhino or elephant) was first evolved from other species, or how a particular baby named John J Johnson conceived by his parents in the last year, a life always emerges from a hierarchy of extremely complex interactions, starting from the lowest level (where the granules are the chemical elements) all the way up along the hierarchy.

An important lesson learned from making a life is emergence is a bottom-up process and top-down is not symmetrical to bottom-up. A top-down analysis of a newborn baby (using a partitioning approach) is the task of anatomy, but this will not reveal how he or she is made. In other words, although top-down lets us examine the product, bottom-up reveals the process. This justifies the need for both top-down and bottom-up directions of research in GrC.

Just like the study of emergence, the study of granular computing is intended to provide a unified approach of studying complex phenomena across various domains. In addition, just like the case of emergence, granular computing models are made concrete in specific application domains. Although emergence is a complex issue and has not caught much attention by researchers in GrC community, certain aspects have been addressed under the

notion of information integration. A more systematic research along this direction is needed. Below we discuss several notable points:

- *Interaction among granules.* The interactions of granules were the beginning of GrC. This consideration marks the differences between rough set theory and granular computing. The issues in knowledge representations, additive measures, belief functions (nonadditive measures), and approximations were discussed in Lin (1998a, b).
- *Relationships of levels within a hierarchy.* In granular computing, granules in a higher level may have greater integrity and higher bond strength than those in a lower level. The structures need to be fully explored to establish a basis of granular computing. However, in the study of emergence, levels in the hierarchy interact in a much more complex way; for example, a completely new form may take place, dramatically different from what can be found at lower level (such as life comes from nonliving chemical materials). In this respect, the relationship of granular computing with mereology should be examined, which is a collection of axiomatic formal systems dealing with parts and their respective wholes and the parts of the parts in a whole.
- *Other important features.* Other important features demonstrated in complex systems, such as self-organization, evolution, adaptation, etc., are not well-supported by current studies in granular computing. However, this situation seems to be changing. For example, Gan et al. (2006) described a self-learning algorithm for uncertain information processing.
- *Building blocks.* Both granular computing and the study of emergence have emphasized the importance of building blocks for model construction but for different purposes.

Granulation Provides a Unified View for Intelligent Problem Solving

Granular computing and related concepts play an important role in a wide range of intelligent

problem solving tasks, including data mining (which is intended to discover implicit, interesting knowledge patterns from massive data) (Han et al. 2012). We now take a look at this issue, which is addressed through three tiers (from the most basic form to the most advanced form), and accordingly, this section is divided into three subsections. We first present how to *interpret* problem domains in terms of concepts in granular computing and then discuss how to *cast* issues of intelligent problem solving in terms of granular computing. Both of these two tiers do not involve any specific algorithms or methods developed in granular computing. Finally, we take on the issue of granular computing algorithms, using granular computing for data mining as an example, which is further illustrated by an implemented case study. Our discussion through these tiers demonstrates that granular computing and its related concepts serve as a unified view for intelligent problem solving in various degrees, even many of the topics examined below are not usually considered as the core of GrC activities.

Re-examination of Existing Studies from a GrC Perspective

Although granular computing should not be a simple restatement of existing results, reexamining or re-interpreting existing result from granularity can help reshape the problem definition and set foot for a new start of granular computing-oriented solutions – so long as our investigation does not stop at the reinterpretation.

Here we use the case of bioinformatics as an example (note this is not intended as a general discussion on the relationship between granular computing and bioinformatics). Various subjects studied in bioinformatics, such as sequence, block, pattern, motif, and protein, can all be viewed as granules at different levels. Below we describe some sample scenarios to illustrate the applicability of GrC in bioinformatics. The key idea behind our discussion is that the relationship between GrC is twofold: GrC can be applied to shed light for bioinformatics problem solving, and bioinformatics also poses interesting challenges for GrC. Materials presented here are mainly taken from resources cited in Chen (2003c).

- *Granulation construction under constraints:* The complexity involved in structural bioinformatics problem solving can take advantage of techniques developed in granular computing and demand new solutions yet to be developed in granular computing. As discussed before, one of the basic concepts in granular computing is granulation. An essential question related to granulation is how to construct higher-level granulation from lower-level granules. As an example, here we examine the fragment assembly problem, which arises when a DNA sequence is broken into small fragments. These fragments must then be assembled to reconstitute the original molecule. (In a sense, the fragments form a covering in the context of GrC). The importance of this problem lies in the fact that with current technology it is impossible to sequence directly contiguous stretches of more than a few hundred bases. On the other hand, there is a technology to cut random pieces of a long DNA molecule and to produce enough copies of the pieces to sequence. Therefore, a typical approach to sequencing long DNA molecules is to sample and then sequence fragments from them. Solving this complex problem consists of assembling a collection of fragments coming from a long, unknown DNA sequence in a correct order and orientation. This can be cast as a problem in GrC, which is concerned with construction of larger granulation from smaller granules *under certain constraints* (such as imposing or preserving a certain order and orientation).
- *Characterization of granulation:* Upon granulations have been constructed, the next question is to uncover the basic features of these granulations (viz., find an *interpretation* of the constructed granulations). As an example, consider the following. Lin (2003) described a data mining algorithm for discovering regions of locus control, i.e., these regions that are instrumental for activating genes. One type of such elements of locus control is the matrix attachment regions (MARs). The discovery of MARs is based on the observation that a group of patterns are bonded together by the virtue of

their similar function. After such a grouping, a search for the patterns in a given group can be performed to identify three regions in the query DNA sequence. If a large subset of members of a functionally related group of patterns is found in a given region of the DNA sequence, one can justifiably classify it as an MAR. The detection problem is to identify observation within a given region. Therefore, discovery of MARs is actually done by grouping for granulation (here in the form of patterns) and then finding the statistic characters of those granulations – somewhat similar to solving a mathematical equation. A more general indication of this approach is the task of studying *characterization of granulations*, namely, what are the significant features of the constructed granulations. Although characterization rule mining has been studied by many researchers in data mining community (Cohen-Solal et al. 2015), the task described in the above example goes beyond the reach of many existing approaches in characterization rule mining, because the granulations considered here may be constructed in a dynamic manner. Therefore, a study in characterization of granulation will enrich both GrC and basic theory for characterization rule mining.

Forming Intelligent Problem Solving Tasks in Terms of Concepts in Granular Computing

Intelligent query processing has very rich contents. GrC (integrated with appropriate AI and data mining techniques) can contribute to intelligent query answering involving aggregate data, mainly due to its power of grouping data in a dynamic and flexible manner. The concepts of constructing equivalent classes, partitioning and covering, as widely used in GrC literature, can be incorporated into the discovery of query-relevant rules from the database. In addition, GrC can be more directly absorbed into presentation techniques applied on original extensional answers to produce final answers in an intensional flavor. Here we take a look on the issue of answering queries at the right granulation levels as originally presented in Chen (2003b).

Although discovery of query-relevant rules is an interesting direction of study, in many other scenarios, we may want to generate aggregate answers from existing extensional answers. Comparing with the previous one, this direction can be considered as a “low key” solution; nevertheless, it deserves equal attention due to its practicality. As the online availability of data is exploding, business firms have begun to use such data to help to *make better decisions* for their business. For example, an auto manufacturer may want to know which type of vehicles made the highest profit in the previous year or which generation of people could be the ideal target population for auto purchase and what are their “tastes.” For such inquiries, a conventional answer, usually extensional answers consisting of all the involved individual tuples, is not the most appropriate answers. Frequently people feel the need for better decision support based on data analysis and knowledge discovery from the primitive data. As a more concrete example, suppose the Board of Directors of a retail firm wants to know something about the salary levels of its employees. A question could be: “In our company, what kind of people is making more than \$100,000 a year?” We may expect to receive several types of answer to this query. An answer could be “All directors and most managers.” It gives some qualitative description about the common characteristics of those qualified individuals. An alternative way is using the aggregate expression. An *aggregate expression* is a sequence of terms that are in the format of “ r/t C ,” where C represents a concept with a total number of t individuals while r is the number of these individuals who belong to the answer. As to the above query, an aggregate answer could be “8/8 directors + 27/32 managers.” From this example it can be seen that an aggregate answer is superior to the qualitative answer “all directors and most managers,” in that the aggregate answer not only covers all the information released in the latter but also provides more quantitative information toward the overall picture of the database. In general, there exists more than one answer to a query. The question is how to obtain the best (optimal) one from those candidate answers.

An approach of using *query entropy* (or Q-entropy for short) for intelligent query answering concerning aggregate answers (with a prototype experiment developed) has been developed for answering aggregate queries at the appropriate data granulation levels (Chen 1999). The particular aspect we are looking for is that, given a conceptual hierarchy C and a query Q against a database D , when multiple answers exist at different granulation levels, how to evaluate them and select the best answer. This approach can be examined from a granular computing perspective. We first note that partition can be viewed as a special kind of granulation. However, in classification rule mining, partition is conducted on the training samples. In contrast, in Q-entropy approach partition is conducted on extensional answers for queries. The approach of using Q-entropy for selecting appropriate levels of query answering also differs from classification rule mining in several other aspects, including size of granules, the construction of the tree, the way in which the entropy is used, and so on. Because of these features, a GrC perspective can be incorporated into presentation techniques applied on original extensional answers and produce final answers in an intensional flavor.

Granular Computing and Data Mining/OLAP

Relationship Between These Two Fields

The interesting relationship between granular computing and data mining can be further examined in many different aspects; here we just take a look at one example on granular operators for property preservation. Granulation allows different representations of the same problem in different levels of detail. It is naturally expected that the same problem must be consistently represented. Granulation and its related computing methods are meaningful only if they preserve certain desired properties. For example, Zhang and Zhang (2003) studied the “false-preserving” property, which states that if a coarse-grained space has no solution for a problem, then the original fine-grained space has no solution. Such a property can be explored to improve the efficiency of problem solving by eliminating a more detailed

study in a coarse-grained space. In the context of hierarchical planning, one may impose similar properties, such as upward solution property, downward solution property, monotonicity, etc. This property resembles a priori principle as discussed in the context of association rule mining (Han et al. 2012) but significantly differs from the view of new properties appearing at higher level of hierarchies which are not found in lower levels, in the study of emergence (Hobbs 1985).

Granular Computing as a Basis for Data Mining and OLAP

Data mining from databases (Han et al. 2012), a key step in the knowledge discovery in databases (KDD), is intended to discover hidden knowledge buried under the ocean of data. As indicated in Lin and Zadeh (2004), “in essence data mining may be viewed as a form of summarization of very large datasets, while granular computing may be viewed as operations on summaries of small datasets. The common rule of summarization in data mining and granular computing is the principle reason why granular computing is of high relevance to data mining.” For example, in Lin (2004), Lin showed that generalized association rules can be expressed as union of basic granules. In fact, an examination of data mining functionalities (or tasks) as discussed in Han et al. (2012) shows a close relationship between the notion of granule and core ideas in each functionality. For example, mining of characterization rules for a concept resembles characterization of a granule, mining of classification rules resembles top-down process of partitioning the universe for all the involved granules, and identifying clusters for a set of granules is corresponding to grouping similar granules into subsets so that a cover of the original set can be established from the union of these subsets. Additional issues on deductive data mining using granular computing have also been discussed.

In a sense, data mining is to find the hidden equivalence relation. For example, cluster analysis is aimed to find clusters, which is a sort of equivalence relation, so at the cluster level, only clusters (rather than original data points) are seen – a typical case of information hiding. Similarly characterization or discrimination analysis

(Cucker and Smale 2002) is to identify dominant features to characterize target class or discriminate target class versus comparison class (with details of attribute values removed) – again a typical example of information hiding. An example of the impact of granulation size to data mining is multilevel association rule mining (Han et al. 2012), which results in revised algorithms and reduced support/confidence level at coarser granulation levels.

Related to data mining is the OLAP (On-Line Analytical Process) technology. Unlike data mining, however, OLAP does not resort to reasoning process; rather, it is more closely tied to database and data warehouse technology for analyzing summary and historical data. Since information granules are formed from data granules, perspectives from GrC also contribute directly to OLAP. In fact, OLAP operations such as roll-up and drill down are directly concerned with granulation sizes. In addition, OLAP techniques have been coupled with data mining, giving GrC more room to maneuver. For example, as discussed in Han et al. (2012), complex aggregations using multifeatured data cubes can facilitate data mining type queries to allow computation of aggregates at different granularity levels. In addition, a unified On-Line Analytical Mining (OLAM) has been proposed (Han et al. 2012). Additional discussion about this integrated data mining/OLAP and the role of GrC can be found in Chen (1999, 2003a) and Zhang and Chen (2002).

Relationship with Soft Computing and Natural Computing

Relationship with soft computing: Probably, the most ancient granule is the intuitive infinitesimals (which refers to the idea of objects so small that there is no way to see them or to measure them). It was the inspiring source of calculus and Zeno’s stationary and moving paradox. A granule is an atom of uncertainty (Lin 2006b, 2008). Yet research related to uncertainty in computational research has focused on the technical side rather than on the philosophical side; therefore, the bulk of research on uncertainty has been conducted in

the realm of soft computing rather than granular computing. Soft computing has been defined as an association of computing methodologies which includes its unique, complementary, and symbiotic constituent members fuzzy logic, neurocomputing, evolutionary computing, and probabilistic computing. Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. In particular, the primary contribution of fuzzy logic is the machinery of granular computing, which serves as a basis for the methodology of computing with words (Zadeh 1998). Suffice it to say that even granular computing and soft computing have significant overlap in contents and in research individuals, and are both driven by human-centered intelligent problem solving, they also differ in the perspectives used in problem solving: unlike granular computing which sets emphasis on granulation-based reasoning under uncertainty and infrastructure of human information processing, soft computing sets emphasis on developing individual algorithms on “soft” side of human information processing which tolerates imprecision and uncertainty.

Relationship with natural computing: Since granular computing is closely related to soft computing and since soft computing is also closely related to natural computing, the relationship between granular computing and natural computing should not be overlooked. Natural computing is the term used to encompass all AI approaches based on some inspiration from nature, soft computing, man-made complex systems, biologically inspired computing, and novel computing paradigms rooted in nature in a higher or lesser degree. However, as noted by Zadeh (1998), in effect, the role model for soft computing is the human mind. There is a significant overlap between natural computing and soft computing in contents; yet they differ in *perspectives* and *emphases*: natural computing is more on how to *model the nature*, while in soft computing more is on how to *achieve* “softness.” A brief examination on the contents of natural computing can reveal the relationship between granular computing and natural computing more directly. Recall that natural computing refers to three types of approaches (De Castro 2006):

computing inspired by nature (such as generic algorithms) which makes use of nature as inspiration for the development of problem solving techniques; *simulation and emulation of natural phenomena in computers*, which is basically a synthetic process aimed at creating patterns, forms, behaviors, and organisms that (do not necessarily) resemble “life-as-we-know-it”; and *computing with natural material*, such as quantum computing. From a GrC perspective, a quantum can be viewed as a granule, and granule is organized in such a way that it could grow exponentially, and then granulation may provide a new framework of attacking classical theory of computing (Granular computing information center). In addition, genetic algorithms can be interpreted as producing new granules from existing granules through specific operations (such as crossover or mutation).

Taking a step further, we note that natural computing is an important field to demonstrate the pervasive concepts of emergence (Hobbs 1985), embodiment (Pfeifer and Bongard 2007), as well as other key concepts such as interactivity, adaptation, learning, evolution, and self-organization. Biologically inspired computing (Bongard 2009) is a new area devoted to these studies. Granular computing can play an important role in the study of natural computing as in the case of learning theory: As noted in granular computing information center, learning is interpolation of data, based on background knowledge (granules). The concepts of neural network, support vector machine, and Smale’s mathematical foundation of learning (Cucker and Smale 2002) can all be formulated along this line of thinking. Granular computing can also enrich itself by learning from natural computing, for example, how to incorporate the notion of emergence to the modeling of human cognitive process, to achieve better human-centered problem solving.

Relationship with Fundamental Issues of Computing and Complex Systems Problem Solving

We now examine a number of more advanced perspectives relating granular computing to basics

of computing and complex systems problem solving.

Granular Computing as Infrastructures for AI-Engineering

As endorsed by Zadeh (2006) at the very beginning, granular computing should serve as foundation of human problem solving. As such, GrC serves as the infrastructure for AI-engineering: uncertainty management, data mining, knowledge engineering, and learning (Lin 2006a). Structures, representations, and applications of granular computing have been discussed in this context (Lin 2003). Note that unlike the human problem solving perspective, this loose definition is computation-driven – even the key notions of granular computing were originally motivated from human-centered information processing.

It has been noted that a simpler and effectively computable view of knowledge is in need. Rough set theory takes a courageous step and assumes that partitions (classifications) are essence of human knowledge. Granular computing takes a softer view: generalized subsets are basic knowledge (Granular computing information center).

This perspective seems serving the de facto standard of “what granular computing is” for majority researchers who are active in this area. For example, Yager (2006) apparently echoes this perspective, where several learning paradigms for granular computing were discussed, although from a fuzzy set perspective.

Note also, according to this definition, there is no urgency calling for a new computer model or AI architecture; we will stay with the von Neumann architecture, as the traditional AI has taken. Nevertheless, proponents of this perspective do relate granular computing to new computational models, such as quantum computing (as already mentioned earlier).

Granular Computing and New Computational Models

Granulation can be interpreted in the context of axiomatic set theory, as shown in Bargiela and Pedrycz (2002, 2006), where information granulation is defined as a *semantically* meaningful grouping of elements based on their indistinguishability,

similarity, proximity, or functionality. The semantics of granules is derived from the domain that has, in general, higher cardinality than the cardinality of the granulated sets. The next question is then how the meaning (semantics) is instilled into real-life information granules. It has been argued that Turing’s computer on its own is unable to respond to external, physical stimuli. To overcome the problem, we may have to advocate the idea of embodiment in AI (Pfeifer and Bongard 2007) and embrace Bain’s approach (2004) of using inherent computational ability of physical phenomena in conjunction with the numerical information processing ability of universal Turing machine (UTM). It has been argued that in a broad context the UTM implementing the virtual computational intelligence function can be referred to as computing with perceptions or computing with words.

The coming of big data presents tremendous challenges and opportunities for granular computing. For example, Zhang et al. (2010) discussed the general issues of how to construct a parallel programming pattern with granular computing and how to perform parallel programming and computing with granules. As a more specific study, Zhang et al. (2010) presented an approach of modeling MapReduce (a well-known framework for parallel distributed computing) using GrC. Since new computational models other than MapReduce are still needed for processing big data, GrC has the potential of making additional contributions. More discussion on GrC and big data is presented in the next section.

Temporal Aspects of Information Granules

Today’s Big Challenges from Big Data

Although no standard definition yet, the term big data usually refers to data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it. Big data is usually characterized by four V’s (volume, velocity, variety, and veracity) (Dong and Srivastava 2013). Therefore, big data implies

a dynamic, ever-changing environment, which always requires integration.

Due to these unconventional features (comparing with data handled by traditional database management systems or information retrieval systems), big data imposes huge challenges and offers tremendous opportunities as well. Toward this end, below we examine how granular computing can help us deal with such challenges, particularly on time-varying data.

Due to its power to intelligent problem solving and information processing, researchers have started applying GrC to deal with challenges from big data, such as parallelizing MapReduce using GrC (Zhang et al. 2011). It is interesting to note that the advent of big data not only provides unprecedented opportunities for GrC to offer solutions on applications in uncharted terrains but also presents extraordinary opportunities to explore some *previously unknown or overlooked features* of granules. An in-depth study of these features not only helps better understanding of information granules but also enhances the status of using GrC for problem solving related to big data. Below we take a look at one of them, along with a newly proposed concept: dynamic granule.

Dynamic Granule: Capturing Dynamic Nature of Granules

Temporal Aspects of Time-Varying Data

Let us start with a brief review on the study of temporal aspects of time-varying data. In fact, time granularities or temporal granularities have long been studied by researchers, including study of temporal granularities in DBMS. For example, a book on time granularities in databases, data mining, and temporal reasoning (Bettini et al. 2000) defined time granularity as a suitable mapping from the set of integers, called index set, to the power set of the time domain. A set-theoretic characterization of time granularity is systematically pursued. The authors elaborate on the given notion of time granularity, identifying the structural properties and relationships between granularities. The key concept is that of *granularity system*. Granularities that come into play in specific contexts or applications are a (usually finite)

subset of those satisfying the given definition. A granularity system formally captures this subset by specifying the time domain, the properties of the member granularities, and their relationships. In addition, the examination on mixing of temporal data at different, user-defined granularities in a single database has been a research focus by different research groups, including Camossi et al. (2009) and Dyreson et al. (2000). For example, in Camossi et al. (2009), it is noted that multiple spatial and temporal granularities are essential to extract significant knowledge from datasets at different levels of detail: they enable to zoom in and zoom out a dataset, enhancing the data modeling flexibility and being instrumental to boost the analysis of information.

Although more in-depth work on temporal aspects from a GrC perspective is yet to be done, recently there have been some interesting developments. A notable study from Maciel et al. (2016) on evolving granular analytics presented a generalized interval evolving possibilistic fuzzy modeling algorithm as an analytics tool capable to process interval data stream and to produce interval forecasts. This approach could be quite useful for analyzing time-varying dynamic data.

Importance of Exploring Dynamic Granules at Big Data Era

Researchers in GrC have noticed the important role of time in studying GrC. For example, Pedrycz used processing and interpretation of time series and granulation of time (treated as a variable) as examples to illustrate the need for granular computing. In fact, an examination of big data reveals that time can do much more: not only as an example of granulation but also an important *feature* of information granule to be considered in big data context. In particular, we are interested in how to deal with uncertainty in time-varying big data due to velocity. An examination of the FBEM approach (Leite et al. 2012) provides useful clue for achieving an abstraction on common features of dynamic data. Other approaches or studies, including behavior mining, deep data, and quantum databases, also offer useful hindsight for developing a theoretical framework of studying such kind of data. As the result of these examinations, an important concept

of *dynamic granule* is proposed, along with an outline of important steps toward the development of a framework involving dynamic granules, for better management of time-varying data at the big data era (Chen 2017).

Other Related Developments

We have examined various aspects within GrC, as well as a wide range of issues related to it. Yet, the human-centered nature of GrC suggests an even broader examination is needed, because the human-centered aspect has also motivated developments of several other research fields, in parallel to GrC.

One such field is cognitive informatics, which is a transdisciplinary expansion of information science that studies computing and information processing problems by using cognitive information processing mechanisms of the brain by using computing and informatics theories (Wang 2003). Yao (2009, 2016) presented a unifying view of cognitive informatics and granular computing, which contains two parts: a high-level understanding of information processing described by a “machines-brain-the abstract” triangle, which is concretized by a special type of information processing with the aid of multiple levels of granularity at the lower level. This cognitive perspective is related to another new research field of brain informatics (Zhong and Chen 2012) which is devoted to the study of the mechanisms underlying the human information processing system (HIPS) in our brain, in an attempt to overcome limitations of current machine intelligence.

Summary

Although still considered as a relatively new research discipline, in the last two decades, granular computing (GrC) has demonstrated matured features of a research area dedicated for human-centered information processing for abstraction, generalization, and unification of granule-based key concepts. We have identified a number of

basic elements serving as the backbone of GrC, such as granular structure, quotient structure, and information hiding. We have also addressed certain issues not being emphasized by some researchers in GrC, such as granule measurement, as well as the relationship between hierarchy and emergence.

We can notice a wide range of theoretical explorations and practical applications of intelligent system developments from *very diverse* perspectives – for example, data mining applications using a rough set approach (or some other approaches). Although such studies on the “periphery” of GrC may eventually contribute to a better understanding of GrC, it is hoped that we can see a more direct exploration and applications on GrC proper, such as developing theorems and algorithms based on the consensus identified above and further applying them for real-world problems, so the vivid discussion on foundations of GrC can turn to fruitful results.

In the main text, while recognizing differences among different researchers, we have tried to present the philosophical foundation of granular computing as a coherent whole. As we are closing this article, we want to point out a few major differences in regard to philosophical foundation observed from various researchers.

As it current stands, granular computing seems to be a loosely structured, somewhat interdisciplinary field: a big umbrella which accommodates researchers in rough set theory, fuzzy set theory, as well as soft computing and data mining – *so long as we all respect granulation as the core of human-centered information processing*. Some researchers may argue that this multidisciplinary nature is not all bad – after all, researches with different backgrounds and different interests can now gather together, exchange their thoughts, and foster new ideas – for example, the increasing publications of hybridization of fuzzy/rough set approaches seem to be a solid proof. Ironically, such kind of successes may also have hampered the development of a unified theory of granular computing, because there seems to be no urgency of bothering it at all.

For those who have a long-term vision of granular computing, viewpoints vary significantly.

Although seemingly everybody agrees granular computing is about human-centered information processing, what is its ultimate objective? As shown in the previous sections, answers could be:

- Assisting human problem solving in new ways of thinking
- Providing infrastructure to AI-engineering
- Serving as the starting point toward new computational models

Our examination has also addressed important issues related to embodiment and emergence. In addition, to cite Zadeh: “In coming years, granular computing is likely to play an increasingly important role in scientific theories-especially in human-centric theories in which human judgment, perception and *emotions* are of pivotal importance” (from Bargiela and Pedrycz (2007), our *Italic*). Therefore, as the study of human-centered information processing, granular computing should also incorporate the role of emotions into its research agenda. The notion of emergence and the entire field of natural computing can shed light to GrC.

Future Directions

In the last 20 years, granular computing has produced a wide range of research fruits and enjoyed increasing publicity. At the time when we celebrate the life of Lotfi Zadeh and his contribution to GrC (along with his many other contributions), we have to ask ourselves: What will GrC be in the next decade and beyond? The answer is stated (or implied) in other articles on granular computing in this encyclopedia and is thus not in the scope of this current article.

As for the research on philosophical foundation of granular computing, what we want to emphasize is that the future of granular computing lies in the development of algorithms from a human-centered problem solving perspective. Yet a continued examination of philosophical foundation will benefit this direction of research. In addition, as noted in Yao (2007), in its first 10 years, granular computing publications has

experienced a linear growth rate; however, the granular computing community has less interaction with other research communities other than fuzzy sets and rough sets. Granular computing community should be aware of this and take steps to promote interdisciplinary research activities. Today, when we examine the history of the second decade of GrC, we have to admit that this observation remains true in certain degree. Although the size of GrC research community continues to grow, efforts are still needed to promote GrC-related concepts in research communities other than GrC itself.

The research agenda of GrC includes a series of methodological and algorithmic issues (Bargiela and Pedrycz 2002), including construction of information granules, characterization of dimension (granularity) of information granules (Singh 2002) (to provide better insight as to the essence of the granulation process and its implications), development of the encoding and decoding mechanisms between levels of hierarchy, research on interoperability which is crucial to the design of systems operating within the realm of various formalisms of information granularity, as well as others.

There is a need for focusing on key ideas behind granular computing yet have been overlooked (or at least not emphasized enough), such as computing with words. In addition, as indicated earlier, important features needed to support complex systems such as self-organization, adaptation, evolution, etc. are currently not well-supported by granular computing research, and this situation should be changed. Although structured thinking and structured problem solving (Bargiela and Pedrycz 2006; Yao 2009, 2016) have received much deserved attention, other features in human cognitive process (including emotion) also deserve attention, because structuredness may emerge from unstructured (or even irrational) thinking, as human beings do.

The future of GrC is also closely tied with big data. Besides dynamicity of granules (to which the study has just begun), other unknown or previously overlooked features of granules should be explored for better understanding on the nature of granules. Such exploration should lead to a more

comprehensive framework for human-centered problem solving, and more effective algorithms can be developed.

Acknowledgment The author thanks Dr. T Y Lin's useful comments for the improvement of an earlier version of this article.

Bibliography

Primary Literature

- Bains S (2004) Physical computation and the design of anticipatory systems. *Proc IEEE SMC*
- Bargiela A, Pedrycz W (2002) Granular computing as an emerging paradigm of information processing (Chap. 1). In: *Granular computing: an introduction*. Kluwer Academic Publishers, Boston
- Bargiela A, Pedrycz W (2006) The roots of granular computing. In: *Proceedings of IEEE conference on granular computing*, pp 806–809
- Bargiela A, Pedrycz W (2007) Toward a theory of granular computing for human-centered information processing. *IEEE Trans Fuzzy Syst*
- Bettini C, Jajodia S, Wang S (2000) *Time granularities in databases, data mining, and temporal reasoning*. Springer, Berlin/Heidelberg
- Bongard J (2009) *Biologically inspired computing*. Computer
- Boyd R, Gasper P, Trout JD (eds) (1991) *The philosophy of science*. Blackwell Publishers, Cambridge, MA
- Burkhardt H, Dufour CA (1991) Part/whole I: history. In: Burkhardt H, Smith B (eds) *Handbook of metaphysics and ontology*. Philosophia Verlag, Muenchen
- Camossi E, Bertolotto M, Bertino E (2009) Spatio-temporal multi-granularity: modelling and implementation challenges
- Chen Z (1999) An integrated architecture for OLAP and data mining, Chapter 6. In: Bramer M (ed) *Knowledge discovery and data mining: theory and practice*. IEE, pp 114–136
- Chen Z (2002) The three dimensions of data mining foundation. In: *Proceedings of ICDM2002 workshop on the foundation of data mining and discovery*, Dec 2002, pp 119–124
- Chen Z (2003a) On granular computing for aggregate data mining. In: *7th international conference on computer science and informatics*, *Proceedings of JCIS*, pp 431–434
- Chen Z (2003b) Discovering rules and answering queries at the right granulation levels. In: *7th international conference on computer science and informatics*, *Proceedings of JCIS*, pp 435–438
- Chen Z (2003c) Bioinformatics as a study of structured granules. In: *International conference on computational intelligence and natural computing*, *Proceedings of JCIS*, pp 1629–1632
- Chen Z (2017) Exploring dynamic granules for time-varying big data. In: *Proceedings of IEEE DataCom*, 2017
- Chidlovskii, B (2001) Schema extraction from XML data: a grammatical inference approach, KRDB'01 workshop (knowledge representation and databases), Rome, 2001. citeseer.ist.psu.edu/chidlovskii01schema.html/
- Cohen-Solal Q, Bouzid M, Niveau A (2015) An algebra of granular temporal relations for qualitative reasoning. <https://hal.inria.fr/GREYC/hal-01189003v2>
- Cotofrei P, Stoffel K (2005) Temporal granular logic for temporal data mining. In: *IEEE international conference on granular computing*, Pekin, pp 417–422. <http://www3.unine.ch/files/content/sites/imi/files/shared/documents/papers/granular.pdf>
- Cucker F, Smale S (2002) On the mathematical foundations of learning. *Bull Am Math Soc* 39:1–49
- De Castro LN (2006) *Fundamentals of natural computing: basic concepts, algorithms, and applications*. Chapman & Hall/CRC
- Dick S, Schenker A, Pedrycz W, Kandel A (2007) Regranulation: a granular algorithm enabling communication between granular worlds. *Inf Sci* 177(2):408–435
- Dong LX, Srivastava D (2013) Big data integration tutorial in *ICDE'13, VLDB'13*
- Dyreson CE, Evans WS, Lin H, Snodgrass RT (2000) Efficiently supporting temporal granularities. *IEEE TKDE* 12(4):568–587
- Gan Q, Wang G, Hu J (2006) A self-learning model based on granular computing. In: *Proceedings of 2006 international conference on granular computing (IEEE GrC 2006)*, pp 530–533
- Granular computing information center. http://www.cs.sjsu.edu/~grc/grcinfo_center/grcinfo_index.php
- Han J, Kamper M, Pei J (2012) *Data mining: concepts and techniques*, 3rd edn. Morgan Kaufmann
- He J (2002) Schema discovery of the semi-structured and hierarchical data. In: *Proceedings of IDEAL (Intelligent Data Engineering and Automated Learning) 2002*. LNCS 2412, pp 129–134
- Hirota K, Pedrycz W (1999) Fuzzy computing for data mining. *Proc IEEE* 87(9):1575–1600
- Hobbs JR (1985) Granularity. In: *Proceedings of the 9th international joint conference on artificial intelligence*, pp 432–435
- Holland JH (1999) *Emergence: from chaos to order*. Perseus Books Group
- Jankowski A, Skowron A (2007) Toward rough-granular computing. In: *Rough sets, fuzzy sets, data mining and granular computing (Proceedings of RSFDGrC 2007)*. LNCS 4482, pp 1–12
- Keet CM (2006) A taxonomy of types of granularity. In: *Proceedings of 2006 international conference on granular computing (IEEE GrC 2006)*, pp 106–111
- Lee TT (1987) An information-theoretic analysis of relational databases – part I: data dependencies and information metric. *IEEE Trans Softw Eng* 13(10):1049–1061

- Leite D, Ballini R, Costa P, Gomide F (2012) Evolving fuzzy granular modeling from nonstationary fuzzy data streams. *Evol Syst* 3:65–79
- Libkin L (2007) Normalization theory for XML. In: Proceedings of XSym, pp 1–13
- Lin TY (1998a) Granular computing on binary relations I: data mining and neighborhood systems. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*, vol 1998. Physica-Verlag, pp 107–121
- Lin TY (1998b) Granular computing on binary relations II: rough set representations and belief functions. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*, vol 1998. Physica-Verlag, pp 121–140
- Lin TY (2003) Granular computing: structures, representations, and applications. In: *Rough sets, fuzzy sets, data mining, and granular computing* (Proceedings of RSFDGrC 2003). LNCS 2639, pp 16–24
- Lin TY (2004) Mining associations by linear inequalities. In: *ICDM 2004*, pp 154–161
- Lin TY (2006a) A roadmap from rough set theory to granular computing. In: *LNAI 4062* (Updated version: Granular computing on coverings: a roadmap from rough set theory to granular computing), revised
- Lin TY (2006b) Granular computing II: infrastructure for AI-engineering. In: *Proceedings of 2006 international conference on granular computing* (IEEE GrC 2006), pp 2–7
- Lin TY (2008) Granular computing: ancient practices, modern theories and future directions. In: Lin TY (ed) *Encyclopedia of complexity and system science: granular computing*
- Lin TY, Zadeh LA (2004) Special issue on granular computing and data mining. *Int J Intell Syst* 19(7):565–566
- Maciel L, Ballini R, Gomide F (2016) Evolving granular analytics for interval time series forecasting. *Granul Comput* 1:213–224
- Pawlak Z (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer
- Pedrycz W. History and development of granular computing. In: *Encyclopedia of life support systems (EOLSS)*
- Pedrycz W, Chen S-M (2016) Editorial. *Granul Comput* 1:93–94
- Pfeifer R, Bongard J (2007) *How the body shapes the way we think: a new view of intelligence*. MIT Press, Cambridge, MA
- Qiu T, Chen X, Huang H, Liu Q (2006) Ontology capture based on granular computing. In: *Proceedings of 6th international conference on intelligent system designs appl (ISDA'06)*, pp 770–774
- Silberschatz A, Korth H, Sudhashan S (2006) *Database system concepts*, 5th edn. McGraw-Hill, Boston
- Singh GB (2002) Discovering matrix attachment regions (MARs) in genomic databases. *SIGKDD Explor* 1(2):39–45
- Skowron A, Stepaniuk J (2001) Information granules: towards foundations of granular computing. *Int J Intell Syst* 16:57–85
- Sowa JF (2000) *Knowledge representation: logical, philosophical, and computational foundations*. China Machine Press
- Wang Y (2003) On cognitive informatics. *Brain Mind Transdiscip J Neurosci Neurophilos* 4:151–167
- Wikipedia. Granular computing (last update 25 Feb 2017). https://en.wikipedia.org/wiki/Granular_computing
- Yager RR (2006) Some learning paradigms for granular computing. In: *Proceedings of 2006 international conference on granular computing* (IEEE GrC 2006), pp 25–29
- Yao J T (2007) A ten-year review of granular computing. In: *Proceedings of 2007 international conference on granular computing* (IEEE GrC 2007)
- Yao Y (2009) Interpreting concept learning in cognitive informatics and granular computing. *IEEE Trans Syst Man Cybern Part B Cybern* 39(4):855–866
- Yao YY (2016) A triarchic theory of granular computing. *Granul Comput* 1(2):145–157
- Yao JT, Vasilakos AV, Pedrycz W (2013) Granular computing: perspectives and challenges. *IEEE Trans Cybern* 43(6)
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) *Advances in fuzzy set theory and application*. North-Holland, Amsterdam, pp 3–18
- Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Comput* 2:23–25
- Zadeh LA (2006) Granular computing – the concept of generalized constraint-based computation. In: *Proceedings of RSCTC 2006*, pp 12–14
- Zadeh LA (2007a) Granular computing – computing with uncertain, imprecise and partially true data. In: *Proceedings of 5th international symposium on spatial data quality* (ISSDQ 2007). http://www.itc.nl/ISSDQ2007/Documents/keynote_Zadeh.pdf
- Zadeh LA (2007b) Granular computing and rough set theory. In: Kryszkiewicz M, Peters JF, Rybinski H, Skowron A (eds) *Rough sets and intelligent systems paradigms* (Proceedings of RSEISP 07), pp 1–4
- Zhang X, Chen Z (2002) Algorithms for finding influential association rules. In: *Proceedings of 1st IEEE international conference on fuzzy systems and knowledge discovery*, pp 499–503
- Zhang L, Zhang B (2003) The quotient space theory of problem solving. LNCS 2639:11–15
- Zhang J, Zhu X, Huang Z (2010) Parallel programming patterns with granular computing. In: *2010 international conference on computer application and system modeling (ICCA SM 2010)*, V7, pp 300–302
- Zhang B, Shi Z, Zhang B (2011) On modeling MapReduce with granular computing. In: *Proceedings of 2011 IEEE international conference on granular computing*, pp 785–789
- Zhong N, Chen J (2012) Constructing a new-style conceptual model of brain data for systematic brain informatics. *IEEE Trans Knowl Data Eng* 24(12):2127–2142
- Zhou G, Liang J (2006) Granular computing model based on ontology. In: *IEEE international conference on granular computing*, pp 321–324

Granular Computing: Practices, Theories, and Future Directions

Tsau-Young Lin
Institute of Data Science, GrC Society, San Jose,
CA, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Classical Examples of Granulation
Formal Models of Granulation
Granular and Related Structures
Integration – A Dual of Granulation
Semantic Views
Future Directions
Bibliography

Glossary

Glossary All terms are explained in classical sets, but implicitly, we are assuming all terms and assertions do include fuzzified versions (if fuzzifiable).

Granulation Granulation is an operation or a process of forming granules, with a granule being a collection of objects (points) that are drawn together by some constraints, such as indistinguishability, similarity or functionality.

Granular structure Granular structure is the collection of granules, in which the internal structure of each granule is visible as a sub-structure. Informally speaking, granular structure is a collection of white box granules.

Quotient structure A quotient structure is the mathematical structure of the collection of granules, in which each granule is regarded as an element (point) of a set, but the interactions among granules are preserved. Informally

speaking, a quotient structure is a collection of black box granules.

The collection of $\{\dots, -2, 0, 2, \dots\}$ and $\{\dots, -3, 1, 3, \dots\}$ is a granular structure. Let E be the first subset (even integers) and O be the second subset (the odd integers). We write the two subsets by $[E]$ and $[O]$, when we think of them as points. Then the collection of $[E]$ and $[O]$ (as points) is the quotient structure.

Neighborhood system (local granular model abb. local GrC model) A domain of interests (a classical set) U is called the universe. To each point p in the universe, a family of subsets is assigned. Such a family (could be empty) and each such subset is called a neighborhood system $NS(p)$ at p and a neighborhood at p , respectively. The collection β of such a family at every point of the universe is called a neighborhood system $NS(U)$ of the universe. Neighborhood and neighborhood system are pre-GrC language; in granular computing, they are called granule and the granular structure, respectively. The pair (U, β) is called a local granular model, since each granule is associated with some points.

Topological neighborhood system A neighborhood system is called a topological neighborhood system, if it satisfies the axioms of topology.

Binary neighborhood system (binary granular model; binary GrC model) A binary neighborhood system is a neighborhood system defined by a binary relation R . A (right) neighborhood is defined as follows: $B(p) = \{x \mid (p, x) \in R\}$. The collection B of $B(p)$ at each p is the (right) binary neighborhood system. Similarly, we can define a left version: A left neighborhood system L is defined by the $L(p) = \{x \mid (x, p) \in R\}$ at every point p . Note that the right and left neighborhood systems determine each other. The pair (U, β) is called a binary granular model, where β is right (left) neighborhood system or R .

pre-Topology Pre-topology is a general term referring to the neighborhood system, which includes the topological neighborhood system and binary neighborhood system as special cases. Technically, it is equivalent to the neighborhood system.

Bag A bag is similar to a set, but allows an element to appear more than once. For example $\{1, 2, 1, 2, 1\}$ is a bag, but not a set. If a bag contains n elements, we may say it is an n -bag. For example, the previous bag is a 5-bag.

Relational structure (relational granular model; relational GrC model) A family of classical sets is called a universe and denoted by $\mathcal{U}$. A Cartesian product of an n -bag of $\mathcal{U}$ is called an n -product set. A n -ary relation is a subset of an n -product set. A collection β of n -ary relations (n could vary) is called a relational structure. The pair (U, β) is called a Relational GrC Model. Note that this relational structure, except the size, is similar to the relational structure (without functions) of the First Order Logic; the Relational GrC Model permits n to be any cardinal number.

Partial covering (global granular model; global GrC model) Let U be a classical set, called the universe. Let $\beta = \{F^1, F^2, \dots\}$ be a family of subsets. Such a β is a partial covering, and (full)covering, if the union of β is the whole universe. The pair (U, β) is called a Global Granular Model (Global GrC Model).

Equivalence Relation A binary relation $\mathcal{R}$ is called an equivalence relation, if it has the following properties: Let u, v , and w be elements of U .

reflexive:

$$u\mathcal{R}u$$

symmetric:

$$u\mathcal{R}v \text{ implies } v\mathcal{R}u$$

transitive:

$$u\mathcal{R}v \text{ and } u\mathcal{R}w \text{ implies } u\mathcal{R}w$$

Partition A partition $\mathcal{P}$ of a classical set U is a collection of subsets that are mutually disjoint and their union is U . Each subset is called an equivalence class. This name is derived from the fact that partition is equivalent to the following equivalence relation: We define $u\mathcal{R}v$, if and only if u and v belong to the same

equivalence class. Such $\mathcal{R}$ is the equivalence relation corresponding to the partition $\mathcal{P}$. Note that a partition is a special type of a granular structure, so an equivalence class is a special granule.

Definition of the Subject

Granular Computing (GrC) is still in its inception stage, we use motivation as its statements of importance.

How Important Is Granular Computing?

Granulation seems to be a natural methodology deeply rooted in human thinking. Many daily “things” are routinely granulated into sub“things”; for example, the human body has been granulated into the head, neck, and so forth. The notion is intrinsically fuzzy, vague, and imprecise. Formalization is difficult, mathematicians idealized/simplified it into the notion of partitions (= equivalence relations), and have developed it into a fundamental part of mathematics, for example, congruence in Euclidean geometry, quotient structures (groups, rings, etc) in algebra, the concept of “a. e.” (almost everywhere) in analysis. Nevertheless, the notion of partitions (see glossary), which absolutely does not permit any overlapping among its granules, seems to be too restrictive for real world problems. Even in natural science, classification *does* permit a small degree of overlapping; there are beings that are both appropriate subjects of zoology and botany. So a more general theory, namely, Granular Computing is needed.

What Is Granular Computing?

It has been a changing paradigm. We believe in incremental developments. Its development may be similar, though not as glorious, to that of classical geometry. Specific geometries, such as, Euclidean, hyperbolic, elliptic geometries, had appeared before the the unified theory was proposed in Klein’s Erlangen program. Except some details, we are on the final line; in GrC2008 keynote, we (this section editor) verbally had proposed to regard the category theory-based model

(Eighth GrC Model) as the final GrC model; see section “Formal Models of Granulation.”

Granular Computing is a recent label coined by Lin and Zadeh (see section “[Introduction](#)”) to denote a set of common, even ancient, concepts and practices. The subject will be presented from various angles: What are the target concept (defined by examples), key constituents, and the current interpretations or semantic views? How far has it been formalized? What are the important applications?

1) The “Definition” of Target Concept: The key concept in Granular Computing is the concept of granulation, which has been “defined” implicitly by a set of intuitive examples (see section “[Classical Examples of Granulation](#)” for more), including Zadeh’s intuitive view. Here is a list of representative examples. In the following, we will call the domain of interests the universe and denote it by U .

1. Many daily things have been routinely granulated into “sub”things; for example, the human body is granulated into head, neck, etc. The notion is intrinsically fuzzy, vague and imprecise. A formal model just had been proposed by this writer in the keynote of GrC2008. Its final form should appear soon in the International Journal of Granular Computing, Rough Sets, and Intelligent Systems.
2. Space and time has been granulated intuitively into infinitesimal granules; a circle was viewed as a polygon with infinitesimal sides. The idea was known to Zeno (490 BC, implicitly in his paradox), Archimedes (287–212 BC) etc. It led to the invention of calculus, topology, and non-standard analysis; we have modeled it in the First GrC Model.
3. The simplest kind of granulation is partition (see glossary). Its algebraic correspondence, equivalence relation, has played an important role in Euclidean Geometry. (300 BC).
4. The Heisenberg uncertainty principle states that, in general, neither the momentum nor the position of a particle can be determined

simultaneously with arbitrary great precision. In other words, a precise measurement of the momentum can only determine a “neighborhood”(granule) of positions and vice versa. The idea is abstract into the Third GrC Model.

5. A collection of fuzzy sets (in fuzzy control) or functions (e. g. Radial-Basis-Functions) which has the universal approximation property is useful granular structure in a function space or a set of fuzzy sets. The idea is modeled in the Sixth GrC Model.
 6. A committee in a human society is a granule in a social network. Observe that each member may play different roles. By viewing the collection of roles as a relational schema, a committee is a tuple, not necessary a subset. The idea is modeled in the Fifth GrC Model and the Second GrC Model.
 7. A mathematical proof or computer program often contains some lemmas or subprograms. These lemmas or subprograms are granules. These are conceptual examples. They are modeled in the Seventh GrC Model for computable domain, and in the Ninth GrC Model for general cases.
 8. Computers or clusters of computers in Grid/Cloud computing are granules. These are hardware examples. This also belongs to the Seventh GrC Model.
 9. Zadeh’s informal definition (Zadeh 1996): “information granulation involves partitioning a class of objects (points) into granules, with a granule being a clump of objects (points) which are drawn together by indistinguishability, similarity or functionality.”
- 2) A Category Theory based Formal model (Eighth GrC Models) is proposed to be *the* Formal Model for GrC. It realizes all classical examples given above; see section “[Formal Models of Granulation](#).” By specifying the abstract category to various ones, eight common models have been explained in this article. Two models are illustrated at the end of this section;

3) The Key Constituents: two operators, three semantic views, and four structures.

1. Two Operators: information granulation and integration. In granulating a problem, a dual action, namely, integrating the solutions of sub-problems is triggered. Here, we highlight some unusual points; see section “[Integration – A Dual of Granulation.](#)”

1. Recursively granulating a problem may take N^p hard time to terminate. Dynamic programming technique has been used in the classical cases.
 2. Two distinct problems may have the same granulation. Divide/granulate and conquer may have deeper meanings. The EXTENSION Functor of Homological Algebra of Commutative group is illustrated. This is an unexplored area.
2. Three Semantic Views: Granules may be interpreted from:
1. Uncertainty Theory: A granule is a unit of lacking precise knowledge.
 2. Knowledge Engineering: A granule is a unit of Basic Knowledge (Information).
 3. How-to-Solve/Compute-it: A granule is a sub-problem or software unit. It is a special type of basic knowledge.

Each view may have its own GrC theory; for example, Concept Approximations are useful in the Second View, while Information Hiding is in Third View; see section “[Semantic Views](#)”.

3. Four Structures: Granular, Quotient, Knowledge and Linguistic Structures:

1. Granular Structure (GrS): It is the collection of all granules. In the case of partition, GrS is the collection of the equivalence classes.
2. Quotient Structure (QS): If each granule is abstracted into a point and the intersections of granules are abstract to the interactions of points, then such a collection of points is called a quotient structure. In the case of partition, the quotient structure is a classical set, called quotient set.

Informally, GrS is a collection of white boxes (the content of granules are

visible), while the quotient set is a collection of black boxes (the contents of granules are hidden). The process of abstracting a granular structure into a quotient structure is called information hiding; see Examples in section “[Granular and Related Structures.](#)”

3. Knowledge structure: By giving each granule (point) in the quotient structure a meaningful symbol then the named quotient structure is called a knowledge structure. The knowledge structure provides an intuitive view of the quotient structure; the symbols and interaction among symbols are in sync with the granules (points) and interactions among granules (points).

In the case of n partitions (equivalence relations), the knowledge structure can be arranged into an n -column relational table. In the case of n binary relations, the table has been called binary information table, granular table or topological table (Lin 1998b; Lin und Liao 2005).

4. Linguistic structure: By giving each granule in the granular structure a word that reflects its meaning. The interactions among these words are reflected implicitly in precisiated natural language (in a knowledge structure, the interactions among symbols are explicitly reflected from the quotient structure). The linguistic structure is the domain of computing with words.

4) Applications to Computer Security, Web Technology, and Complex Data.

1. Discretionary Access Control Model (DAC) (Lin 1989a, 2003). This structure has been captured in the Third GrC Model. On the basis of this, information flows on DAC can be analyzed; this has been considered a very “difficult” area. As a consequence the Aggressive Chinese Wall Security Policy can be enforced, namely, the system can guarantee that a company’s data will never flow into “enemy” hands,

where “enemy” is a granule of companies that are in conflict.

- Documents can be clustered into a simplicial complex (a common structure in Combinatorial Topology (Spanier 1966)) of keywords and co-occurring keyword sets. The set of keywords can be regarded as a set of vertices, and the collection of co-occurring keywords (within a small neighborhood) is a set of simplexes. Together, they form a simplicial complex (Lin und Chiang 2005; Lin et al. 2006). Simplexes are granules; a simplicial complex is a Second GrC model.
- Granular computing has been used to solve the modeling problem of complex architectures (Liu et al. 2008). It uses the Fifth GrC model.

Illustration of Working Formal Models

The most general model is expressed in the category theory. However, for easiness, we explain the Second GrC model first.

1) Second GrC Model Let U be a classical set, called the universe. Let $\beta = \{F^1, F^2, \dots\}$ be a family of subsets. Then the pair (U, β) , called the Global GrC Model or the Second GrC Model. The β , some time, is called Partial Covering (PCov).

A *granule* can be intuitively defined as a clump of objects that are drawn together by the constraints X , where X can be indistinguishability, similarity or functionality and etc. (Lin 1998a; Zadeh 1996). In the Second GrC model, a granule is a set, namely, we have implicitly assumed that the constraints are uniform. In general, each object may receive distinct constraints, so in the Fifth GrC model, a granule is a tuple, not necessarily a set: One can regard the constraints as a schema (of a relational database), and a granule is a tuple under such a schema. In this case the collection of granules are tuples from various relations.

To understand the next model, Table 1, that explains the generalization process, may be helpful.

Granular Computing: Practices, Theories, and Future Directions, Table 1 Generalize second to fifth GrC Models

second GrC	Generalized to	5th GrC
U	$\rightarrow$	$\mathcal{U} = \{U_j^h, h, j, = 1, 2, \dots\}$
$F^1 \subseteq U$	$\rightarrow$	$R^1 \subseteq U_1^1 \times U_2^1 \times \dots$
$\dots$	$\dots$	$\dots$
$F^j \subseteq U$	$\rightarrow$	$R^j \subseteq U_1^j \times U_2^j \times \dots$
$\dots$	$\dots$	$\dots$

2) Fifth GrC Model

- Let $\mathcal{U} = \{U_j^h, h, j, = 1, 2, \dots\}$ be a given family of classical sets, called the universe. Note that distinct indices do not imply that the sets are distinct.
- Let $U_1^j \times U_2^j \times \dots; j = 1, 2, \dots$ be a family of Cartesian products of various lengths.
- Recall that an n -ary relation is a subset $R^j \subseteq U_1^j \times U_2^j \times \dots \times U_n^j$.
- Let $\beta = \{R^1, R^2, \dots\}$ be a given family of n -ary relations for various n .

Then the pair $(\mathcal{U}, \beta)$, called the Relational GrC Model, is the formal definition of the Fifth GrC Model.

Introduction

What is Granular Computing (GrC)? The best approach is to trace how the intuitions have been evolved. For this purpose, let us recall the event. In the academic year 1996–97, when Lin had his sabbatical leave at Berkeley, Zadeh suggested granular mathematics (GrM) to be his research area. To limit the scope, Lin proposed the term granular computing (Zadeh 1998). So, at the beginning, roughly GrC is the computable part of granular mathematics.

What is Granular Mathematics (GrM)? Zadeh in his 1979 paper (Zadeh 1979) had implicitly explained his view. Here, we take a simpler view: It is a new mathematics, in which “points” are replaced by or associated with “granules.” We

call this process granulation, and the collection of granules the granular structure.

What is a Granule? This is the main topic. There is an obvious candidate, namely, an equivalence class of a *partition* that is an ancient notion in mathematics. In general, a granule can be a crisp/fuzzy subset, a function, an algorithm, a random variable (measurable function), a generalized functions etc.

Traditionally, “how to solve it” (Polya 1957) cannot be any part of formal mathematics, however, “how to compute it” is an integral part of computing. So GrC has included the mathematical/computational problem solving practices.

Classical Examples of Granulation

Ancient Examples

E1 Granulation of the Human Body: The Granules Are Head, Neck, Body, Hand etc.

Many daily things are routinely granulated into “sub”things, probably since ancient time. Human body is granulated into head, neck, etc. Currently, there are no flawless formal models to capture such intuitive concept yet. The obvious one does not work well: One can easily write down a fuzzy membership function to represent a head, a neck or a body. However, the likelihood of any two persons to write down the same membership function for the same granule (e. g., head) is extremely low. Hence we need a much more subtle theory. A new proposal on qualitative fuzzy set theory is in preparation by this author, this example can be realized.

E2 Granulation of Space and Time The space and time has been granulated intuitively into infinitesimal granules by early scientists; this notion was known to Zeno, Archimedes, etc. This intuitive notion led to the invention of calculus by Newton and Leibniz. However, its formalizations, theory of limit (eighteenth century), topology (early twentieth century (Munkres 2000)) and nonstandard analysis (mid twentieth century (Robinson 1966)) are relatively recent. This ancient example, inspired two models, Local GrC model (First GrC Model) and Global GrC Model (Second GrC Model).

E3 Partition The simplest kind of granulation is partition (see glossary). Its algebraic correspondence, equivalence relation, has played an important role in the Euclidean Geometry (300 BC).

Modern Examples

E4 Local Granules of Uncertainty – From Quantum Mechanics

Heisenberg’s uncertainty principle states that, in general, neither the momentum nor the position of a particle can be determined simultaneously with arbitrary great precision. In other words, a point of the momentum (a precise measurement) can determine only a “neighborhood” of positions and vice versa. The idea is simplified into the Third GrC Model (Binary GrC Model).

E5 Local Granules of Basic Knowledge – Discretionary Access Control Models in Computer Security

This example is a common data structure in many computer systems. For each user, there is a set of users (friends) who can access his files, or a set of users (foe), who cannot access his files; this set is called explicitly denied access list.

Examples E4 and E5 are “serious” examples. The idea is modeled in the third GrC Model. Mathematically, it is equivalent to a binary relation. Geometrically, a binary relation is a graph, or network.

E6 Global Granules of Knowledge – Simplicial Complexes

A simplicial complex consists of two objects: One is a finite set of vertices, another one is a family of subsets, called simplexes, of vertices that satisfy the closed condition, namely a subset of a simplex is a simplex. It is a common mathematical structure in combinatorial topology. Currently, it is finding its way to web technology (Lin und Chiang 2005).

Further Examples

Next, we give examples of non-commutative granules, which is a generalization of a binary relation (Binary GrC Model; Third GrC Model).

E7 A Committee in a Human Society Is a Granule in a Social Network A committee in a

human society (a set of human beings) is a granule. Observe that each member may play different roles, so the committee may not consist of homogeneous members; so the members cannot exchange their roles. By viewing the collection of roles as a relational schema, a committee is a tuple. This idea is modeled in the Fifth GrC Model. Observe that different types of committees have different schema. The set of committees under the same schema forms a relation. A collection of n -nary relations for various n in a society can be viewed as a granulation of the society.

In these examples, granules can be fuzzy sets. Since a fuzzy set is characterized by a membership function, so we have generalized the idea further.

E8 A Granule Can Be a Function, a Random Variable (A Measurable Function) or Even a Generalized Function (Such as the Dirac Delta Function) This class of examples led to the Sixth GrC model.

Traditionally, “how to solve it” (Polya 1957) has not been any part of formal mathematics, however, “how to compute it” is an integral part of computing. So GrC includes some mathematical/computational practices.

E9 A Granule Can Be a Lemma in a Mathematical Proof, a Subprogram in a Program, or a Machine/Cluster of Machines in a Grid/Clouding Computing Environment Formally, within the computable domains, a granule is a sub-Turing machine. This is modeled in the Seventh GrC Model. For general cases, they are included in the Ninth GrC Model.

E10 Computers or Clusters of Computers in Grid/Cloud Computing Are Granules These are hardware examples. This also belongs to the Seventh GrC Model.

Formal Models of Granulation

On the basis of these examples, nine models are discussed. The category theory-based model (Eighth GrC Model) has been proposed by this writer in the keynote of GrC2008 as *the formal*

model of GrC. The rest of the eight models are basically “convenient models” in the sense that they can be derived from the most general model, but for convenience, they are modeled independently.

First GrC Models and Ancient Examples

This model is derived from the Ancient Example E2, which probably is the most lively example in the early notion of granules. It led to the invention of calculus by Newton and Leibniz. Actually the idea was much more ancient; it was in the mind of Archimedes, Zeno, etc. Yet the solutions were in modern time. Two formalizations had emerged. One was the concept of limit (nineteenth century) that led to the notion of topology (early twentieth century). The other one was the nonstandard analysis (mid twentieth century), which formally realized the original intuition.

These developments conclude that the following concepts are “equivalent” (not in the formal mathematical sense, as the first one is merely an intuitive notion):

1. The ancient intuitive notion of infinitesimal granule.
2. The formal infinitesimal granule in the non standard analysis.
3. Topological Neighborhood System (TNS) in the standard world.

In other words, the ancient intuition of infinitesimal granules (with the required properties) is realized, not by a set, but by a family of subsets, that satisfies the axioms of topology. Nevertheless, in this paper a (modern) granule will refer to a neighborhood, but, not to the whole neighborhood system.

The notion of topology can be defined in two ways:

1. A topology τ is a family of subsets, called open sets, that satisfies the (global version) axioms of topology.
2. A topology, called topological neighborhood system (TNS), is an assignment that associates to each point p a family of subsets, $TNS(p)$, that satisfies the (local version) axioms of topology.

These two definitions lead us to First and Second GrC Models (Local and Global GrC Models). In this subsection, we will focus on the First GrC Model: Let U and V be two classical sets. Let NS be a mapping, called neighborhood system (NS),

$$NS : V \rightarrow 2^{P(U)},$$

where $P(X)$ is the family of all crisp/fuzzy subsets of X . 2^Y is the family of all crisp subsets of Y , where $Y = P(U)$. In other words, NS associates each point p in V , a family $NS(p)$ of crisp/fuzzy subsets of U . Such a subset is called a neighborhood (granule) at p , and $NS(p)$ is called a neighborhood system at p .

Definition 1 (First GrC Model) The 3-tuple (V, U, β) is called the **Local GrC Model**, where β is a neighborhood system (NS). If $V = U$, the 3-tuple is reduced to a pair (U, β) . In addition, if we require NS to satisfy the topological axioms, then it becomes a TNS.

Some Intuition Behind the NS The following arguments are adopted from my pre-GrC paper (Lin 1997)

(1) The Meaning of “Near”.

The notion of near is rather difficult to formalize. Let us examine the following two examples.

1. Is Santa Monica “near” Los Angeles? Answers could vary. For local residents, who have cars, answers are often “yes.” For visitors, who have no cars, answers may be “no.”
2. Is 1.73 “near” $\sqrt{3}$? Again answers vary; it depends on what should be the appropriate tolerance radius.

Intrinsically “near” is a subjective judgment. One might wonder whether there is a scientific theory for such subjective judgments? Mathematicians have offered a nice solution. They simply include all contexts into its formalism. Here is the formalism of the second question: Given the

radius of an acceptable error, say, radius of errors 1/100 (a given context)

Is 1.73 “near” $\sqrt{3}$?

With the agreement 1/100 is acceptable, then 1.73 is near $\sqrt{3}$! In this case, all possible contexts are ε that represents all positive real numbers. Similarly, if a neighborhood system has been assigned to each city in the Los-Angeles-area. For example, based on car driving, public transportation, walking etc., we assign a neighborhood to *each* city for *each* context. Under such a concept of a neighborhood system, we could have a definite answer for Example 1. So a proper formulation for such a question is:

Assuming that we are taking public transportation (a given context), Is p near q ?

Therefore, a neighborhood system is a good infrastructure for addressing the concept of “near”! These analysis leads to the following conclusions.

1. In Modeling, a neighborhood system is a good infrastructure for providing all possible contexts.
2. Under this model, in an application, selecting a context means selecting a *fixed* neighborhood as a unit of tolerance (uncertainty).

Now, under this concept, we will re-examine previous examples.

Example 1 If we have chosen “driving half an hour” as acceptable distance, then Santa Monica is “near” Los Angeles.

Example 2 Let the collection of ε -neighborhoods be the neighborhood system for the real numbers R ; then $(R, \varepsilon\text{-neighborhoods})$ is a First GrC Model, where ε could take any real value. Now, we re-state the previous example using this First GrC Model

1. Assuming we have agreed $\varepsilon = 1/100$ is acceptable, then 1.73 is “near” $\sqrt{3}$.

2. But, if we have only agreed $\varepsilon = 1/1000$ then 1.73 is not “near” $\sqrt{3}$.
3. Next let us consider a deeper question:

Is the sequence $1, 1/2, 1/3, \dots, 1/n, \dots$ “near” zero?

Then, it is possible, we can have a “yes” answer for all contexts: For any given context, namely, $\varepsilon > 0$, there is a number $N = [1/\varepsilon] + 1$, such that, for all $n > N$, $1/n$ is “near” zero, where $[1/\varepsilon]$ denotes the biggest integer $\leq 1/\varepsilon$.

Readers who are familiar with the standard (ε, δ) -definition of limit can spot the origin of neighborhood systems. Such a context-free (all possible contexts) answer is precisely the classical notion of limits, $\lim_{n \rightarrow \infty} 1/n = 0$. Using our language, we may say that limit is the context free answers of “near”.

Perhaps we should also point out here that there are no context free answers for the question whether two points are “near.”

Brief pre-GrC Historical Notes

1. In 1988–1989, Lin generalized TNS to the Neighborhood Systems (NS) by simply dropping the (local version) axioms of topology (Lin 1988, 1989b) and apply it to approximate retrievals. Each neighborhood was treated as a unit of uncertainty.
2. In the same year (1989), Lin also examined a non-reflexive and symmetric binary relation (conflict of interests) for computer security from the view of NS (Lin 1989a).
3. Abstractly, Lin imposed the NS structure on the attribute domains for approximate retrieval. Taking this view, we should mention that earlier D. Hsiao imposed equivalence relations on the domain for access precision in early 1970 (Hsiao und Harary 1970; Wong und Chiang 1971). In 1980, S. Ginsburg and R. Hull had imposed partial ordering on attribute domains (Ginsburg und Hull 1981; Ginsburg und Hull 1983).
4. In much earlier, NS was studied in (Sierpinski 1952) as a generalization of topology. Note that however, there are fundamental differences, for example, the concept of closures are different. The term pre-topology also has been used for referring NS and TNS.

5. In the early GrC period, Lin, by mapping the NS onto Zadeh’s intuitive definition, used NS as his first mathematical GrC model (Lin 1998a, b, 1999).

Second GrC Models and Modern Examples

As in the previous case, by dropping the global axioms of topology, we have the Second GrC model.

Definition 2 (Second GrC Model) The Pair (U, β) , where β is a family of subsets of U , is called **Global GrC Model**. The β , some time, is referred to as a partial covering (PCov).

Note that the Second GrC model is a special case of the First GrC model: If we regard the sub-collection of all members of the partial covering β , that contains p , as a neighborhood system at p , then this Second GrC model is an example of the First GrC model.

The modern example, simplicial complexes, is an important example of such a model: A simplicial complex consists of a set of vertices and a family of subsets, called simplexes, that satisfies the closed condition (Spanier 1966).

[Digression] Perhaps, it is worthwhile to note that

- The closed condition of the simplicial complex is the a priori principle in association (rules) mining.

This observation plays an important role in document clustering (Lin et al. 2006).

Third and Fourth GrC Models and Modern Examples

In this section, we will build a new model that realizes modern example E4 and E5. Recall that E4 concludes that a precise measure of the momentum can only determine a (probabilistic) “neighborhood” of positions; and E5 concludes that in computer security, the Discretionary Access Control Model (DAC) assigns to each user p a family of users, $Y_p, i = 1, \dots$, who can access p ’s data. In other words, each p is assigned a granule of friends.

To formalize these examples, let U and V be two classical sets. Each $p \in V$ is assigned a subset, $B(p)$, of “basic knowledge” (a set of friends or a “neighborhood” of positions).

$$p \rightarrow B(p) = \{Y_i, i = 1, \dots\} \subseteq U.$$

Such a set $B(p)$ is called a (right) binary neighborhood and the collection $\{B(p) \mid \forall p \in V\}$ is called the binary neighborhood system (BNS).

Definition 3 (Third GrC Model) The 3-tuple (U, V, β) , where β is a BNS, is called a **Binary GrC Model**. If $U = V$, then the 3-tuple is reduced to a pair (U, β) .

Observe that BNS is equivalent to a binary relation (BR):

$$\text{BR} = \{(p, Y) \mid Y \in B(p) \quad \text{and} \quad p \in V\}.$$

Conversely, a binary relation defines a (right) BNS as follows:

$$p \rightarrow B(p) = \{Y \mid (p, Y) \in \text{BR}\}.$$

So, both modern examples give rise to BNS which was called a binary granular structure in (Lin 1998a). We would like to note that based on this (right) BNS, the (left) BNS can also be defined:

$$D(p) = \{Y \mid p \in B(Y)\} \quad \text{for all} \quad p \in V\}.$$

Note that BNS is a special case of NS, namely, it is the case when the collection $\text{NS}(p)$ is a singleton $B(p)$. So the Third GrC Model is a special case of the First GrC Model.

The algebraic notion, binary relations, in computer science, is often represented geometrically as graphs, networks, forests etc. So the Third GrC Model has captured most of the mathematical structure in computer science.

Next, instead of a single binary relation, we consider the case where β is a set of binary relations. It was called a [binary] knowledge base (Lin 1998a). Such a collection naturally defines an NS.

Definition 4 (Fourth GrC Model) The Pair (U, β) , where β is a set of binary relations, is called

Multi-Binary GrC Model. This model is most useful in data bases; hence it has been called Binary Granular Data Model (BGDM), in the case of equivalence relations, it is called Granular Data Model (GDM).

Observe that a Fourth GrC Model can be converted, say by a mapping $\mathcal{G}$, to a First Model. Conversely, a First GrC Model induces, say by $\mathcal{F}$, to a Fourth Model. So First and Fourth models are equivalent, but not naturally, namely, $\mathcal{G}$ and $\mathcal{F}$ are not inverse to each other.

Models for Further Examples

We have observed in section “[Definition of the Subject](#)” that the collection of n objects that are “drawn together” is, not necessary a subset, but is a tuple in an n -ary relation. For example, if the universe is a human society then a group of people may be drawn into a committee with distinct roles, such as the chair, vice chair, secretary, treasurer, etc. As every member has a different role, they can not be swapped around. So the committee is not a set; it is a tuple under the schema that consists of distinct roles.

Definition 5 (Fifth GrC Model)

1. Let $\mathcal{U} = \{U_j^h, h, j = 1, 2, \dots\}$ be a given family of classical sets, called the universe. Note that distinct indices do not imply that the sets are distinct.
2. Let $U_1^j \times U_2^j \times \dots$ be a family of Cartesian products of various length.
3. Recall that an n -ary relation is a subset $R^j \subseteq U_1^j \times U_2^j \times \dots \times U_n^j$.
4. Let $\beta = \{R^1, R^2, \dots\}$ be a given family of n -ary relations for various n .

The pair $(\mathcal{U}, \beta)$, called the Relational GrC Model, is a formal definition of the Fifth GrC Model.

Note that this granular structure is the relational structure (without functions) in the First Order Logic, if n only varies through finite cardinal number.

For the next two models, we will use the language of category theory in next sub-section. We

may note that we have not committed ourselves to every specific details yet.

Definition 6 The Sixth GrC Model is in the categories of functions, random variables, and even generalized functions.

Fuzzy sets are described by membership functions, so granules can be regarded as membership functions; note that the First to Fifth GrC Models include fuzzy sets. Hence, we consider further generalizations: granules are functions, random variables (measurable functions), generalized functions (e. g. Dirac delta functions).

In the case, a granule is a function, we may require that the granular structure (the collection of granules) has the universal approximation property, namely, any function in the universe can be approximated by the functions in the collections. The membership functions selected in fuzzy controls do have such properties. In neural networks, the functions generated by the activation functions also have such a property (Park und Sandberg 1991).

In the case of probability/measure theory, quantum mechanics may be a good guiding example.

Definition 7 The Seventh GrC Model is in the category of Turing machines.

For example, a collection of lemmas in a mathematical proof (mechanizable), a set of subprograms in a computer program, or a computer or cluster of computers in grid/cloud computing are granules in the model.

Definition 8 The Ninth GrC Model is in the category of qualitative fuzzy sets.

This model was proposed after the Eighth GrC model. It has not been published in printing form yet. The idea is similar to the model that we have called it sofset (this is not a typo) (Lin 1996). It associates to each “real world” fuzzy set, a collection of membership functions; please watch for new development.

Category Theory Based Models

Now we generalize the category of sets to general categories. It is somewhat a surprise that this is basically the same as the category of relational databases (Lin 1990). In other words, the abstract structures of data and knowledge are similar. After analysis, it seems reasonable; because in GrC approach, the basic unit of knowledge is a granule of data.

Let us set up some language for the Category Theory. A category consists of

1. A class of objects, and.
2. A set $\text{Mor}(X, Y)$ of morphisms for every ordered pair of objects X and Y , which satisfies certain properties. For this paper, the formal details are not important; we only need the language loosely.

Here are some examples.

1. The Category of Sets: The objects are classical sets. The morphisms are the maps.
2. The Category of Sets with binary relations as morphisms: The objects are classical sets. The morphisms are binary relations. This is the Category of Entity Relationship Models.
3. The Category of Power Sets: The object U_X is the power set $P(X)$ of a classical set X . Let U_Y be another object, where Y is another classical set. The morphisms are the maps, $P(f) : U_X \rightarrow U_Y$ that are induced by maps $f : X \rightarrow Y$.

Let CAT be a given category.

Definition 9 (Category Theory Based GrC Model)

1. $C = \{C_j^h, h, j, = 1, 2, \dots\}$ is a family of objects in the category CAT.
2. There are families (which are bags; see the glossary) of Cartesian products, $C_1^j \times C_2^j \times \dots$ of objects, $j = 1, 2, \dots$ of various lengths. They are called product objects.
3. An n -ary relation object R^j is a sub-object of the product object $C_1^j \times C_2^j \times \dots \times C_n^j$.

4. $\beta = \{R^1, R^2, \dots\}$ be a family of n -ary relations (n could vary).

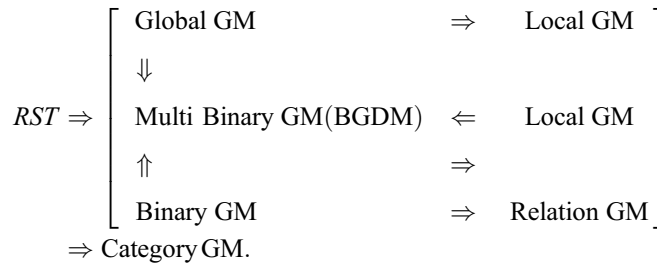
The pair (C, β) , called the Categorical GrC Model (Eighth GrC model), is the formal model of granulation.

By specifying the general category to various special cases, we have all models: By specifying the category to be the category of sets, we have the Fifth GrC model. Further by limiting n to 2, we have the First GrC Model and the Fourth GrC Model. By assuming the symmetry for all n -ary relations, we have the Second GrC Model. By restricting the number of relations to be one and

$n = 2$, we have the Third GrC Model. For the Sixth and the Seventh, further researches are needed.

Overview of Early GrC Models

Schematically we summarize the relationships based on the Granular Structures of early GrC Models as follows: (the diagram will be different, if it is based on their approximation spaces). “ $\Rightarrow \Leftarrow$ ” is a two-way generalization but they are not inverse to each other. “ $\Rightarrow$, $\Uparrow$, and $\Downarrow$ ” are one way generalizations. “GM” means GrC Models and RST means Rough Set Model.



Granular and Related Structures

Four structures are related to each other, they are all briefly explained in section “[Definition of the Subject](#).” We will focus on a quotient structure here.

A granule can be examined from three states: We will use the following example to illustrate the idea. Let U be the set $(Z, +)$ of the integers as an additive group, let E and O be the set of even and odd integers respectively.

1. Isolated State (Internal State): Let us consider the case E is in the isolated state, that is, independent from the universe U . In this case, only the internal structure of E is available to us: we can only know that the number 2 is playing the role of the identity in an additive group, but is *not* aware of the existence of 1 in the outside of E . So in isolated state, E is the additive group $(Z, +)$. Some time, this state may also be called

internal state as only the internal structure is involved.

2. Embedded State (Conceptual State): In this state, E is a subgroup of U . We not only know that the number 2 is playing the role of the identity in E and we also *do know* that 2 is an even integer in U .

[Digress] In category theory, a subobject (in this case a subgroup) is often represented by a pair (a group, the mapping of the group into the subgroup),

$$(Z, +) \xrightarrow{\text{homomorphism}} E \subseteq (Z, +).$$

The pair (the first $(Z, +)$, $\xrightarrow{\text{homomorphism}}$) is the subobject (subgroup). The first component $(Z, +)$ represents the internal structure of (or isolated state of) even integers, the middle E is the embedded state of even integers. Since this state involves with every aspect of the concept of a granule, one may call it conceptual state.

3. Quotient State (External State): Note that the quotient structure Q consists of two elements, namely, the two subset of even and odd integers (E and O , respectively). But the two subsets appear as two elements, $[E]$, and $[O]$ in Q . Note the bracketed symbols emphatically denote their roles as points (elements) of a set; no contents of subsets are visible. The interactions between the two subsets induce interactions between these two elements. So Q is the integer mod 2 (as additive group). In notations,

$$Q = (\mathbb{Z}_2, +) \equiv \{[0]_2, [1]_2\}$$

as an additive group. So the quotient state of E is the element $[0]_2$ in Q . Since the quotient state only involves the external relationships with other granules, this state may also be called external state.

With these preparations, we can define the concepts of granular structure (GrS), quotient structure (QS) and knowledge structures (KS): GrS is the collection of granules in embedded states. The quotient structure is the structure of granules in quotient states, namely, each granule is abstracted to an element (point) and the interactions among granules are abstracted to the interactions among elements (points).

Among these four structures only the quotient structure is a difficult concept. Here, we will illustrate only the quotient structure. We will start from the simplest case, namely, the GrS is a partition. By definition, a partition is a collection of equivalence classes that are mutually disjoint and their union is the whole universe. Hence, by abstracting each equivalence class to a point, we have a set of points that have no interactions. This implies:

Proposition 1 *The quotient structure of a partition is a classical set.*

Next, we need a lemma.

Lemma 1 *The collection of complete inverse image of a given mapping $f: U \rightarrow V$ defines a partition. Namely, the collection $\{f^{-1}(f(p)) \mid p \in U\}$ forms a partition.*

So NS and BNS induce partitions on U . These partitions have been called the derived partition (or derived equivalence relation); see (Lin 1998a).

Let $B: U \rightarrow P(U)$ and $NS: U \rightarrow 2^{P(U)}$ be a binary neighborhood system (BNS) and neighborhood system (NS), respectively. Then (U, NS) and (U, B) are First and Third GrC Models, respectively. By the previous Lemma, the complete inverse images of the two mappings, B and NS , are partitions (equivalence relations). The two partitions, denoted by E_B and E_{NS} , induce two quotient sets U/E_B , U/E_{NS} , respectively. Moreover, the pre-topologies, BNS and NS, of U naturally induce the respective pre-topologies on the quotient sets U/E_B and U/E_{NS} , respectively (Lin 1989b, 1992).

Proposition 2 *Such pre-topological spaces U/E_B (BNS-space) and U/E_{NS} (NS-space) are the quotient structures of Binary and Local GrC Models (Third and First GrC Models) respectively.*

This example indicates that there are additional pre-topological structures on the information tables (Chiang et al. 2005; Lin und Liao 2005).

Let (U, β) be a Global GrC Model, where $\beta = \{F^1, F^2, \dots\}$ be a partial covering. We will use $[F^{j_1}], [F^{j_2}], \dots$ to denote a set of points (when we think of $F^1, F^2, \dots$ as points).

Proposition 3 *The granular structure, β , generates a semi-group $S(\beta)$ under set theoretical intersection. Then the quotient structure of $(U, S(\beta))$ is a semi-group generated by $[F^{j_1}], [F^{j_2}], \dots$ under "intersection" $[F^{j_1}]^\circ [F^{j_2}] = [F^{j_1} \cap F^{j_2}]$.*

This example indicates that there are additional algebraic structures on the information tables (Lin 2006b).

We observe that each quotient structure is not easy to determine. It depends on what are the mathematical structures under consideration. Let us recall some works from pure mathematics. Let the universe U be the ring Z of integers. In ring theory, the collection β of all prime ideals is the center of attentions. Let p be a prime number, then the prime ideal consists of following set

$$\{\dots, -2p, -p, 0, p, 2p, \dots\},$$

together with some algebraic structure.

By regarding each prime ideal as a point, the collection of prime ideals is a set, often denoted by $\text{Spec}(Z)$ in algebraic geometry. The structure of prime ideals turns $\text{Spec}(Z)$ into a topological space under Zariski topology.

Example 3 (Two Granular and Quotient Structures)

1. The quotient structure of (U, β) , where β is the collection of the prime ideals, is a topological space $\text{Spec}(Z)$ (Mumford 1988).
2. The quotient structure of $(U, S(\beta))$, where $S(\beta)$ is the semi-group generated by the intersection, is isomorphic to the semigroup of positive integers.

In RST, the approximation spaces of the first and second examples are different. However in GrC, the approximations (under the knowledge engineering view) of the two examples are the same. In GrC, granules represent known basic units of knowledge (known concepts), hence the intersections of known concepts are known concepts. So in GrC we take all possible intersections of granules to approximate unknown concepts. This example clearly indicates the effectiveness of this view.

Integration – A Dual of Granulation

Classical Divide (partitioning) and Conquer is extended to Granulate and Conquer. In other words, we are considering the cases, in which the sub-problems may not be logically independent from each other. So the standard thinking may not work, for example

- It may take Np-hard time to granulate a problem, from top to bottom.

This issue is not new. Some cases have been addressed in the topic of “dynamic programming” in data structure courses. We have used

“topological divide” to solve the cases in the Third GrC model (Lin 2005).

In this section, we will explore the following problem: For convenience, we will use the term, sub-structure, to denote (1) the collection of the internal structures of each granule and (2) the quotient structure. Now, we raise the following questions:

- Can we find more than one universes (granular models) that have the same sub-structure?
- Equivalently: Could two distinct problems, after granulation, have the same sub-structure?

In homological algebra, this is called the extension problem; in computer science, we feel that integration is a better name.

We will illustrate the concept in the following two cases

- 1) Simple Integration: no information structure.
- 2) Integration with information structure.

Simple Examples of Integration

This is a truly unexplored area; we will explain the case when granulations are partitions.

A Simple Partition We will illustrate the key concepts by simple examples. Let the universe U be the set of all integers, $Z = \{\dots, 1, 0, 1, \dots\}$ with the following granular structure, namely, a given collection of two granules

$$\{\{\dots, -2, 0, 2, \dots\}, \{\dots, -3, 1, 3, \dots\}\}.$$

We name the granules E and O (the even and odd integers). Using the languages of section “Granular and Related Structures,” we say E and O are in the embedded state. We will use $[E]$ and $[O]$ to denote the quotient state. So the set $Q = \{[E], [O]\}$ is the quotient set. Intuitively, Q is a set of black boxes; the internals of black boxes are invisible.

Now we will summarize the important structures of U so that integration can be formulated:

Let $\text{Int}(X)$ denote the internal structure of a granule X , that is, the structure of X in isolation (and not a subset of U).

1. $A = \text{Int}(\{\dots, -2, 0, 2, \dots\})$ is the set of integers Z .
2. $B = \text{Int}(\{\dots, -3, 1, 3, \dots\})$ is the set of integers Z .
3. Two copies of integers, A and B , are mapped to even integers and odd integers respectively.
4. Q is the quotient set that consists of $[E]$ and $[O]$ as two elements (points).

Schematically, we give the following situation:

$$Z = \left\{ \begin{array}{l} A = \text{Int}(E) \rightarrow E \\ B = \text{Int}(O) \rightarrow O \end{array} \right\} (\subseteq) U \rightarrow Q = \{[E], [O]\}.$$

Integrations on Partitions From the point of view of problem solving,

- The quotient structure represents the recipe (a set of higher level instructions) of integrating the sub-solutions. It is the “Main Program” that integrates the returns of sub-program calls.

Here a sub-solution means the solution of a sub-problem that has been solved in isolation. Now the integration problem can be formulated as follows:

1. We only know that two copies of integers, A and B , are mapped to the two subsets of unknown universe.
2. The images of the mappings form a partition in the unknown universe U .
3. But we do know its quotient structure $Q = \{[E], [O]\}$, namely a set of two elements, where $[E]$ represents the point that is abstracted from the image E of A in U ; similarly $[O]$ is that of B .

Schematically, we can summarize it as follows: The unknown U has two unknown subsets, image A and image B , that form a partition on unknown U , but with a known quotient set which is equivalent to the integers mod 2.

$$Z = \left\{ \begin{array}{l} A \rightarrow \text{Image } A \\ B \rightarrow \text{Image } B \end{array} \right\} (\subseteq) \text{Unknown } U \rightarrow Q = Z_2$$

Can We Construct the Unknown Universe? And Is It Unique? The answer is “yes” and it is unique. It is the Cartesian product $Q \times Z$. Observe that from a set-theoretical point of view, $Q \times Z$ is equivalent to Z , so the constructed one is the old friend.

Integration on Partition with Information Structure Let us consider a second view on the set of integers. But this time, the universe carries additional information, namely, the additive structure of integers, $(Z, +)$. This universe is denoted by $(U, +)$. Then

1. $\text{Int}(E)$ is the *additive group* $(Z, +)$ of integers, and $\text{Int}(O)$ is a *set* Z of integers.
2. The quotient structure $(Q, +) = \{[E], [O]; +\}$ is an additive group: $[E] + [E] = [E]$, $[E] + [O] = [O] + [E] = [O]$, $[O] + [O] = [E]$. This $(Q, +)$ is often called the integer mod 2, and denoted by $(Z_2, +)$.

Again, we give a similar situation

$$\left\{ \begin{array}{ll} (Z, +) = \text{Int}(E) & \xrightarrow{(\text{homomorphism})} E \\ Z = \text{Int}(O) & \xrightarrow{(\text{map})} O \end{array} \right\} \subseteq (U, +) \rightarrow (Q, +).$$

The upper arrow

$$(Z, +) = \text{Int}(E) \xrightarrow{(\text{homomorphism})} (U, +) \rightarrow (Q, +)$$

is very similar to a short exact sequence (SES) in homological algebra.

Now, we will extend the SES to granules:

$$\text{Int}(\text{GrS}) \xrightarrow{\text{map}} U \rightarrow QS,$$

where (1) $\text{Int}(\text{GrS})$ is the collection of the internal structures of all granules, (2) its image in U (under the map $\rightarrow_{\text{map}}$) is GrS , the granular structure in U , and (3) QS denote the quotient structure of GrS . The map, $\rightarrow_{\text{map}}$, is on-to-one within each granule,

but two granules may have overlapping images in U . We will call this a *granular exact sequence*.

Can $(U, +)$ be reconstructed back?

The answer is more than yes; there are *two* solutions! The two solutions are: $(Z_2 \times Z, +)$ and $(Z, +)$. They are not equivalent as additive groups. This fact can be expressed by the extension functor, namely, $EXT(Z, Z_2) \neq 0$.

The important question is

Could such a functor be extended to formal granular models?

The answer may be “yes.” The granular exact sequences shall play a similar role as short exact sequences. There are some preliminary results in (Lin 2006a); for example, the fact that the knowledge representation of a symmetric binary relation is complete, implies that the integration is unique.

Integrations on Non-Partition Case We will consider a simple example using the concept of granular exact sequence. Let $U = \{a, b, c\}$, let β be the set of all subsets of U . So the internal structure of the granules are: (we will use n -granule to denote a granule of n elements) One copy of 3-granule; three copies of 2-granule, and three copies of 1-granule. This is a Global GrC Model. The quotient structure is a collection of 7 elements (to find the quotient structure is not easy, but easy to verify). These elements are related by partial ordering, called “a face of.” We will use the geometry to represent this quotient structure; it is a triangle spanned by $\mathbf{i}(1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, $\mathbf{k} = (0, 0, 1)$. The triangle can be viewed geometrically (a simplicial complex of closed triangle) as ONE open triangle, THREE open segments, and THREE vertices. Its a partial ordered set of 7 elements.

Now the granular exact sequence (three steps) can be described as follows:

1. 3-granule $\{g_1^3, g_2^3, g_3^3\} \rightarrow$ a subset of unknown universe $\rightarrow$ open triangle $\Delta(\mathbf{ijk})$
2. first 2-granule $\{g_1^2, g_2^2\} \rightarrow$ a subset of unknown universe $\rightarrow$ first open segment $\Delta(\mathbf{ij})$,

3. second 2-granule $\{g_3^2, g_4^2\} \rightarrow$ a subset of unknown universe $\rightarrow$ second open segment $\Delta(\mathbf{ik})$,
4. third 2-granule $\{g_5^2, g_6^2\} \rightarrow$ a subset of unknown universe $\rightarrow$ third open segment $\Delta(\mathbf{jk})$,
5. first 1-granule $\{g_1^1\} \rightarrow$ a subset of unknown universe $\rightarrow$ first vertex $\Delta(\mathbf{i})$,
6. second 1-granule $\{g_2^1\} \rightarrow$ a subset of unknown universe $\rightarrow$ second vertex $\Delta(\mathbf{j})$,
7. third 1-granule $\{g_3^1\} \rightarrow$ a subset of unknown universe $\rightarrow$ third vertex $\Delta(\mathbf{k})$.

We will use $I(g_j^i)$ to denote the image in the unknown universe. We do know many distinct g_j^i mapped to the same point in the quotient structure. Here are the mappings

1. 3-granule $\{I(g_1^3), I(g_2^3), I(g_3^3)\} \rightarrow \Delta(\mathbf{ijk})$
2. first 2-granule $\{I(g_1^2), I(g_2^2)\} \rightarrow \Delta(\mathbf{ij})$,
3. second 2-granule $\{I(g_3^2), I(g_4^2)\} \rightarrow \Delta(\mathbf{ik})$,
4. third 2-granule $\{I(g_5^2), I(g_6^2)\} \rightarrow \Delta(\mathbf{jk})$,
5. first 1-granule $\{I(g_1^1)\} \rightarrow \Delta(\mathbf{i})$,
6. second 1-granule $\{I(g_2^1)\} \rightarrow \Delta(\mathbf{j})$,
7. third 1-granule $\{I(g_3^1)\} \rightarrow \Delta(\mathbf{k})$.

From the maps above, we may make some identifications: Observe that from the maps, first, second, fifth, we can identify, without losing generality, that $I(g_1^3) = I(g_1^2) = I(g_1^1)$. By similar arguments, we have

1. $I(g_1^3) = I(g_1^2) = I(g_3^2) = I(g_1^1) \rightarrow \Delta(\mathbf{i})$
2. $I(g_2^3) = I(g_2^2) = I(g_5^2) = I(g_2^1) \rightarrow \Delta(\mathbf{j})$
3. $I(g_3^3) = I(g_4^2) = I(g_6^2) = I(g_3^1) \rightarrow \Delta(\mathbf{k})$

So the 12 points are actually three points; they will be denoted by $\{I(g_1^3), I(g_2^3), I(g_3^3)\}$. Thus, the granular structure of the unknown universe is: one 3-granule $\{I(g_1^3), I(g_2^3), I(g_3^3)\}$, three of its 2-subgranules $\{I(g_1^3), I(g_2^3)\}$, $\{I(g_1^3), I(g_3^3)\}$, $\{I(g_2^3), I(g_3^3)\}$ and three of 1-subgranules $\{I(g_1^3)\}$, $\{I(g_2^3)\}$, $\{I(g_3^3)\}$. So we have re-captured the unknown U and β . – This is

the integration. The key question is: Is this U unique? The answer is “no”; we will skip it here.

Semantic Views

Granules may be interpreted from three views.

1. **Uncertainty Theory:** A granule is a unit of lacking precise knowledge. Both L. A. Zadeh and T. Y. Lin, who coined the label, started from the uncertainty theory. Lin took his neighborhood system as a system of uncertainty. Zadeh has a grand project (Zadeh 1998).
2. **Knowledge Engineering:** A granule is a unit of basic knowledge (Information). In their book, D. Stanat and D. McAllister state “Knowledge varies in sophistication from simple classification to ...” (Stanat und McAllister 1977). A classification is a partition, so an equivalence class is a (unit of) basic knowledge; this view is also promoted by Pawlak (1991). So we believe: a granule is a (unit of) basic knowledge.
3. **How-to-Solve/Compute-it:** A granule is a subproblem or a software unit. It is a special type of basic knowledge.

Each view may have its own GrC theory: Some fundamental operators are:

- 1) *Information Hiding:* It is a transformation of granular structures into quotient structures (see glossary).

A quotient structure is the mathematical structure of the collection of granules, in which each granule is regarded as an element (point), and the interactions among granules are transformed into the interactions among elements. For example, in group theory a quotient group is a collection of cosets, in which each coset is regarded as an element and the multiplication of cosets is abstracted into the multiplication of elements in the quotient group. Using the software engineering language, a granule in a quotient structure is a black box, while in a granular structure, it is a white box. These are easy cases, for granulations

that have nonempty overlappings, the quotient structure may not be easy to determine; see section “Granular and Related Structures”.

- 2) *Information Integration:* see section “Integration – A Dual of Granulation.”
- 3) *Knowledge Representation:* This amounts to give each point in the quotient structure (section “Granular and Related Structures”) a meaningful name; we will call it naming map. So knowledge representation is a composition of information hiding (a map from a granular structure to a quotient structure) and the naming map.

Knowledge representation is another way to discover new knowledge by organizing the knowledge structure in an appropriate fashion, such as, an information table (= a relation in database theory). In GrC, the table often has additional algebraic and/or topological structures (Lin 2005; Lin und Liao 2005).

- 4) *Concept Approximation:* Among three semantic views, the knowledge engineering view is the most suitable view for this operation. Under this view, concept approximation is to express approximately the unknown concepts (arbitrary subsets of the universe) in terms of known basic knowledges (granules).

It is reasonable to regard that “and” ($\cap$) or “or” ($\cup$) operations of two basic units of known knowledge are also a known knowledge. So, we take finite intersection and any number of union as acceptable knowledge operations; the last one is derived from the topological spaces. However, we believe a negation of a known knowledge is not necessary a piece of known knowledge; so negation is not an acceptable operation.

Note that this approximation is different from *Rough Set Approximations*, which, including its generalizations, are based on the sole operation “or” ($\cup$). In other words, RST does not regard the “and” of two known concepts as also a known concept. So strictly speaking, rough set approximation is not a concept approximation.

GrC has many models; each model has a slightly different approximation theory. First, we will explain.

The *Second GrC Model (Global GrC Model)*: Let C_1 be a given collection. Let $\mathcal{G}_1$ be the collection of all possible finite intersections of C_1 . Then, by definition, the pair (U, β) is a Second GrC Model (Global GrC Model), where $\beta = C_1$. Let G be a variable that varies through the collection $\mathcal{G}_1$, then we define.

Definition 10 Three approximations

1. Upper approximation: $C[X] = \overline{\beta}[X] = \{p : \forall G, \text{ such that, } p \in G \ \& \ G \cap X \neq \emptyset\}$.
2. Lower approximation: $I[X] = \underline{\beta}[X] = \{p : \exists a \ G, \text{ such that, } p \in G \ \& \ G \subseteq X\}$.
3. Closed set-based upper approximation: (Sierpinski 1952) used *closed* closure operator. It applies closure operator repeatedly (transfinitely many steps) until the results stop growing. The space is called Frechet(V)-space or (V)-space. $Cl[X] = X \cup C[X] \cup C[C[X]] \cup C[C[C[X]]] \dots$ (transfinite). For such a closure, it is a closed set.

Theorem 1 *The concept approximation space of the Global GrC Model (Second GrC Model) is a topological space.*

We should also note that under the rough set approximation, this model is not a topological space.

Next, let us consider the approximation theory of.

The *First GrC Model (Local GrC Model)*: Before, we proceed, let us examine the classical case, the topological space (U, τ) . A subset $N(p) \subseteq U$ is a neighborhood of p , if $N(p)$ contains an open set that contains p . The union of all such open sets is the interior points of $N(p)$, which is the largest open set in $N(p)$; let us denote it by $O(p)$. Observe that $O(p)$ consists of every point that regards $N(p)$ as its neighborhood.

Now, we will generalize this idea to the First GrC Model. Let $NS(p)$ be the neighborhood system at p . Let $G(p)$ be the collection of all finite

intersections of all neighborhoods in $NS(p)$. Let G be a variable that varies through $G(p)$.

Definition 11 With such a G , the previous equations given above define the appropriate notions of $C[X]$, $I[X]$, $Cl[X]$ for the First, Third and Fourth GrC Models.

We should caution here that in the Third GrC Model, there is at most one neighborhood at each point, so there is no “true” intersection.

Let $N(p)$ represent an arbitrary neighborhood of $NS(p)$. Let $C_N(p)$, called the center set of $N(p)$, consist of all those points that have $N(p)$ as its neighborhood. (Note that $C_N(p)$ is the generalization of $O(p)$ in TNS).

Now we will observe some harder question: Do the intersections of neighborhoods at distinct points belong to some $G(p)$?

Proposition 4 (Theorem of Intersections in NS)

1. $N(p) \cap N(q)$ is in $G(p) = G(q)$, iff $C_N(p) \cap C_N(q) \neq \emptyset$.
2. $N(p) \cap N(q)$ is not in any $G(p) \ \forall \ p$, iff $C_N(p) \cap C_N(q) = \emptyset$.

If we regard $N(p)$ as a known basic knowledge, then we should define the knowledge operations: Let $\circ$ be the “and” operation of the basic knowledge (a neighborhood). For technical reasons, the $\emptyset$ is regarded as a piece of the given basic knowledge.

Definition 12 (New Operations in NS)

1. $N(p) \circ N(q) = N(p) \cap N(q)$, iff $C_N(p) \cap C_N(q) \neq \emptyset$.
2. $N(p) \circ N(q) = \emptyset$, iff $C_N(p) \cap C_N(q) = \emptyset$.

Observe that BNS is a special cases of NS. So we have:

Definition 13 Let B be a BNS, then

1. $B(p) \circ B(q) = B(p) = B(q)$, iff $C_B(p) = C_B(q)$.
2. $B(p) \circ B(q) = \emptyset$, iff $C_B(p) \cap C_B(q) = \emptyset$. Note that $B(p) \cap B(q)$ may not be empty, but it is not a neighborhood of any point.

Observe that in Binary GrC Model, two basic knowledges are either the same or the set theoretical intersection does not represent any basic known knowledge.

5) Higher Order Concept Approximations.

In the Fifth GrC model, we consider the relations (subsets of product space) as basic knowledge. Any subset in a product space is a new unknown concept. We will illustrate the idea in the following case: U^j is either a copy of V or U . Moreover, in each product space, there is at most one copy of V , but no restrictions on the number of copies of U . If a Cartesian product has no V component, it is called a U -product space. If there is one and only one copy of V , it is called a product space with unique V .

1. u and u^1 are said to be directly related, if u and u^1 are in the same tuple (of a relation in β), where $u \in U$ and u^1 could be an element of U or V .
2. u and u^2 is said to be indirectly related, if there is a finite sequence $u_i, i = 1, 2, \dots, t$ such that (1) u_i and u_{i+1} are directly related for every i , and (2) $u = u_1$ and $u^2 = u_t$.
3. An element $u \in U$ is said to be v -related ($v \in V$), if u and v are directly or indirectly related.
4. v -neighborhood, U_v , consists of all the $u \in U$ that are v -related.

In such a relational GrC model (U^j with unique V) induces a map:

$$B: V \rightarrow 2^U; v \rightarrow U_v.$$

Such a map defines a binary neighborhood system (BNS), where U_v is a v -neighborhood in U , and hence induces a *binary GrC model* ((U, V, B)). Next, we will consider the case $U = V$ and define.

Definition 14 The high order approximations of the Fifth GrC model are the approximations based on the v -neighborhood system.

6) Concept Approximations in other Categories.

In a category of functions, we will be interested in those granular structures that have the universal approximation property. For a category of Turing machines (algorithms), it is still unclear how to define the concept approximations.

Future Directions

Granular Computing is still in its inception stage; possible directions are wide open. Here we will focus only on those issues that are touched in this article.

1) Developments of Categories.

In this paper, a category based model is proposed as *the* Formal Model for GrC. It can be specialized into various models to realize all the classical examples, including the first example, the granulation of the human body. We should note that the claim on the realization of the first example is not in printing form yet. However, the author feels that it is important to inform the readers that is occurring.

The key to realize the first example is based on the category of qualitative fuzzy sets or softsets (this is not a typo); please watch for new development. The categories of functions, random variables (measurable functions) and Turing machines must be developed, too.

2) Developments of Granular Structures.

Given a granular structure, we associate it with four structures (including itself). Among them, quotient and knowledge structures are mathematical consequences of a granular structure (if it is given mathematically). However, the linguistic structure is not a mathematical formalism but is a natural language formulation. In this paper, there is no report on this direction. We urge the readers to read Zadeh's article.

3) Imported Concepts.

For information integration (this may correspond to Zadeh's term, "organization"), we have illustrated the idea imported from homological algebra. It is unclear if we have

imported the correct thinking; but it does point out essential problems in granulate and conquer.

4) GrC and RST.

RST has been served as the “model” of GrC developments. So, even there are a lot of similarities here, we would like to caution the readers that there are fundamental differences. For example, the fundamental views of uncertainty are quite different; Pawlak used “unable to specify” as the base of uncertainty, while GrC regards a granule as a unit of uncertainty (such as uncertainty in quantum mechanics). Also the approximation theories are different. Of course, there are other differences; we skip.

5) GrC, Databases and Data Mining.

As we have pointed out that the categorical structures of databases and GrC are similar; at the same time, we need to point out the differences in semantics. Nevertheless, we are looking forward to the transfer of database technology to GrC. For data mining, please see the database section on the articles by this author on deductive data mining using GrC, and mining decision rules using RST.

6) GrC and Fuzzy Logic.

Most of the expositions have been based on classical sets (and fuzzifiable concepts). For more intrinsic fuzzy view, we strongly recommend the readers to read Zadeh’s article.

7) GrC and Clouding Computing.

Theoretically, cloud computing can be related to the GrC on the category of Turing machines. We expect some strong interactions in near future.

As we have observed that GrC is deeply rooted in human thinking, we expect GrC will have many interactions with wide variety of areas.

Bibliography

- Baliga P, Lin TY (2005) Kolmogorov complexity based automata modeling for intrusion detection. In: 2005 IEEE international conference on granular computing, Beijing, 25–27 July 2005. pp. 387–392
- Chiang IJ, Lin TY, Liu Y (2005) Table representations of granulations revisited. In: Proceedings 10th international conference, RSFD-GrC 2005, Regina, Canada, 31 Aug–3 Sept 2005, part I. Lecture notes in computer science, vol 3641. Springer, Berlin, pp 728–737
- Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: Slowinski R (ed) Intelligent decision support: handbook of applications and advances of the rough sets theory. Kluwer, Dordrecht, pp 203–232
- Ginsburg S, Hull R (1981) Ordered attribute domains in the relational model. In: XP2 workshop on relational database theory, June 22–24, Pennsylvania State University, USA
- Ginsburg S, Hull R (1983) Order dependency in the relational model. *Theor Comput Sci* 26:149–195
- Hsiao D, Harary F (1970) A formal system for information retrieval from files. *Commun ACM* 13(2):67–73. (Corrigenda 13(4))
- Lin TY (1988) Neighborhood systems and relational database. In: Proceedings of the 16th ACM annual conference on computer science, Atlanta, 23–25 Feb 1988. p 725
- Lin TY (1989a) Chinese wall security policy – an aggressive model. In: Proc of the 5th aerospace computer security application conference, Tuscon, 4–8 December, pp 286–293
- Lin TY (1989b) Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings of the 4th international symposium on methodologies of intelligent systems (poster session), Charlotte, 12 Oct 1989. Pp 75–86. (Available to the public from the National Technical Information Services, U. S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161. NTIS price codes rited copy: A11 Micorfiche A01. Available to DOE and DOE contractors from the office of scientific and technical information P. O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576–8401)
- Lin TY (1990) Relational data models and category theory (abstract). In: CSC’90, Proceedings of the ACM 18th annual computer science conference on cooperation, Sheraton Washington Hotel, Washington DC, 20–22 Feb 1990. p 424
- Lin TY (1992) Topological and fuzzy rough sets. In: Slowinski R (ed) Decision support by experience – application of the rough sets theory. Kluwer, Dordrecht, pp 287–304
- Lin TY (1996) A set theory for soft computing. In: Proceedings of 1996 IEEE international conference on fuzzy systems, New Orleans, 8–11 Sept 1996, pp 1140–1146
- Lin TY (1997) Neighborhood systems – a qualitative theory for fuzzy and rough sets. In: Wang P (ed) Advances in machine intelligence and soft computing, vol IV. Duke University, Durham, pp 132–155
- Lin TY (1998a) Granular computing on binary relations I: data mining and neighborhood systems. In: Skoworn A Polkowski L (eds) Rough sets in knowledge discovery. Physica, Heidelberg, pp. 107–121

- Lin TY (1998b) Granular computing on binary relations II: rough set representations and belief functions. In: Skowron A, Polkowski L (eds) *Rough sets in knowledge discovery*. Physica, Heidelberg, pp 121–140
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh L, Kacprzyk J (eds) *Computing with words in information/intelligent systems*. Physica, Heidelberg, pp 183–200
- Lin TY (2003) Chinese wall security policy models: information flows and confining trojan horses. In: De Capitani di Vimercati S, Ray I, Ray I (eds) *Data and applications security XVII: status and prospects*, IFIP TC-11 WG 11.3 17th annual working conference on data and application security, Estes Park, 4–6 Aug 2003. Kluwer, Boston, pp 275–287
- Lin TY (2005) Divide and conquer in granular computing: topological partition. In: *Proceedings of the 2005 North American fuzzy information processing society annual conference: computing for real world applications*, Ann Arbor, 22–25 June 2005
- Lin TY (2006a) A roadmap from rough set theory to granular computing. In: *Proceedings of the 1st international conference rough sets and knowledge technology*, RSKT 2006, Chongqing, 24–26 July 2006. *Lecture notes in computer science*, vol 4062. Springer, Berlin, pp. 33–41
- Lin TY (2006b) Granular computing on partitions, coverings, and neighborhood systems. *J Nanchang Inst Technol* 25(2):1–7. (Special issue as the Proc of the international forum on theory of GrC from rough set perspective)
- Lin TY, Chiang IJ (2005) A simplicial complex, a hypergraph, structure in the latent semantic space of document clustering. *Int J Approx Reason* 40(1–2):55–80
- Lin TY, Liao CJ (2005) Granular computing and rough sets. In: Maimon O, Rokach L (eds) *The data mining and knowledge discovery handbook*. Springer, New York, pp 535–561
- Lin TY, Huang KJ, Liu Q, Chen W (1990) Rough sets, neighborhood systems and approximation. In: *Proceedings of the 5th international symposium on methodologies of intelligent systems, selected papers*, Knoxville, 25–27 Oct 1990, pp 130–141. Library of Congress Catalog Card Number:90–84690
- Lin TY, Sutojo A, Hsu JD (2006) Concept analysis and web clustering using combinatorial topology. In: *Workshops proceedings of the 6th IEEE international conference on data mining (ICDM 2006)*, Hong Kong, 18–22 Dec 2006, pp. 412–416
- Liu Y, Lin TY, Xu C, Zhang Q, Huang L, He P (2008) Modeling complex architectures based on granular computing on ontology. Submitted to *IEEE Trans Fuzzy Syst*
- MacLane S (1995) *Homology*. Classics in mathematics, reprint of the 1975 edn. Springer, Berlin, x+422 pp
- Mumford D (1988) *Introduction of algebraic geometry*. In: *The red book of varieties and schemes*, mimeographed notes from the Harvard mathematics department 1967, *Lecture Notes in Mathematics*, vol 1348. Springer, Berlin
- Munkres J (2000) *Topology*, 2nd edn. Prentice-Hall
- Park JW, Sandberg IW (1991) Universal approximation using radial-basis-function networks. *Neural Comput* 3:246–257
- Pawlak Z (1991) *Rough sets*. Theoretical aspects of reasoning about data. Kluwer, Dordrecht
- Polya G (1957) *How to solve it*, 2nd edn. Princeton University Press, Princeton
- Robinson A (1966) *Non-standard analysis*. North-Holland, Amsterdam
- Sierpinski W (trans by Krieger C) (1952) *General topology*. Mathematical exposition, no 7. University of Toronto Press, Toronto
- Spanier EH (1966) *Algebraic topology*. Springer, New York
- Stanat D, McAllister D (1977) *Discrete mathematics in computer science*. Prentice-Hall, Englewood Cliffs
- Wong E, Chiang TC (1971) Canonical structure in attribute based file organization. *Commun ACM* 14(9):593–597
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) *Advances in fuzzy set theory and applications*. North-Holland, Amsterdam, pp 3–18
- Zadeh LA (1996) The key roles of information granulation and fuzzy logic in human reasoning. In: *1996 IEEE international conference on fuzzy systems*, New Orleans, 8–11 Sept, p 1
- Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 90:111–127
- Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Comput* 2:23–25
- Zimmerman H (1991) *Fuzzy set theory – and its applications*. Kluwer, Dordrecht

Principles and Perspectives of Granular Computing

Jianchao Han¹ and Nick Cercone²

¹Department of Computer Science, California State University Dominguez Hills, Carson, CA, USA

²Faculty of Science and Engineering, York University, Toronto, ON, Canada

Article Outline

Glossary

Definition of the Subject of Its Importance

Introduction

Perspectives of Granular Computing

Neighborhood System: A Mathematical Structure of Granular Computing

Rough Set System: An Equivalence Neighborhood System

Fuzzy Set System: A Fuzzy Neighborhood System

Granular Computing in Data Mining

Granular Computing in Software Engineering

Future Directions

Bibliography

Glossary

Granular Computing Granular computing is a strategy of problem-solving. The basic idea of granular computing comes from the strategy of divide-and-conquer. It is a twofold process: first, it granulates a complex program into granules, and then it computes these granules and integrates results to form a solution to the complex problem. Granulation of problems into granules is of different forms such as chunking, clustering, partitioning, division, or decomposition, while granules are clumps of objects or points. Computations with granules

are either within granules or granule with environment.

Neighborhood System Neighborhood system is a mathematical structure of granular computing to model granules and can be used to compute structure of granules and/or between granules and ambient spaces. A neighborhood system at a point is a framework to capture the concept of “near” objects, and any subset of objects can be approximated by a set of neighborhoods. A neighborhood system defines a set of binary relations, and a set of binary relationships can be used to define a neighborhood system.

Fuzzy Set Theory Unlike the classic set theory where a set is represented as an indicator function to specify if an object belongs or not to it, a fuzzy set is an extension of a classic set where a subset is represented as a membership function to characterize the degree that an object belongs to it. The indicator function of a classic set takes value of 1 or 0, whereas the membership function of a fuzzy set takes value between 1 and 0.

Rough Set Theory The rough set theory deals with inexact information systems. In an information system, a decision table consists of a set of objects which are characterized by a set of condition attributes and decision attributes. Objects in the decision table can be classified into equivalence classes using an *indiscernibility* relation, and equivalence classes are explored to approximate crisp sets of objects.

Data Mining Data mining is a very important step of knowledge discovery in databases to extract nontrivial, previously unknown, and potentially useful patterns that are hidden from large data sets. Data mining tasks include classification and clustering analysis of objects into categories, discovery of associations and correlations among data items, characterization and summarization of subsets of objects,

finding sequential patterns and similarities in ordered data, etc.

Software Engineering Software engineering is an engineering discipline that applies a systematic approach to produce reliable and efficient software. Software development process consists of different phases, including requirements, analysis, design, specification, implementation, testing, deployment, and maintenance. Object-oriented methodology to software engineering is based on several techniques and principles such as inheritance, modularity, polymorphism, and encapsulation.

Definition of the Subject of Its Importance

Granular computing (abbreviated, GrC) is a general computing paradigm that effectively deals with elements and granules that are generated from elements. Granules are vaguely viewed as generalized subsets of the universe of discourse and may be in different forms such as classes, clusters, groups, and intervals. The essence of GrC is to build an efficient computational model for complex applications with huge amounts of data, information, and knowledge. Though the term has been proposed recently, the basic ideas and principles of granular computing have been explored in a variety of fields under the different names such as divide-and-conquer and have been extensively applied in many problem-solving disciplines such as politics, sociology, economics, and computational intelligence with different forms either consciously or unconsciously by human beings.

Research on GrC will not only reveal the computational structures of complex problems and systems, build the mathematical foundations of complex problem-solving, and formalize the common problem-solving strategy but also help practitioners explore novel specific problem-solving strategies in special application domains. Additionally, by consciously using GrC strategies and heuristics, human being has a better chance to find a good solution to a complex problem.

Introduction

The terminology of granular computing (GrC) was first proposed by Professor T. Y. Lin in 1996 as a label of family of theories, methodologies, and techniques in computational intelligence that make use of granules, although this strategy can be dated back to the ancient time. Its basic ideas and principles have been studied in various application domains. Especially in the form of partitions, the theory has been accumulated for hundred of years in mathematics.

Before the term, granular computing, was invented, some results and thoughts on granular computing had been achieved. The explicit study of granular computing is originated from Zadeh. In 1979, Zadeh (1979) introduced the notion of information granulation and suggested that fuzzy set theory might find potential applications in this regard.

Though the granular computing is focusing on the non-partition theories, nevertheless, the partition case was the main source of inspiration, especially the partition theory in computer sciences. In 1982, Pawlak (1998) proposed rough set theory (RST) to deal with inexact information. It is an uncertainty theory using a very special form of granules, called equivalence classes. It is primarily due to the rough set theory (partition theory) that causes researchers to realize the importance of generalized granular computing notion.

In 1985, Hobbes (1985) presented a theory of granularity as the base of knowledge representation, abstraction, heuristic search, and reasoning. In his theory the problem world is represented as various grains, and only interesting ones are abstracted to learn concepts. The conceptualization of the world can be performed at different granularities and switched between granularities. Even though his discussions mainly concentrated on the partition cases, his model is more general than rough sets and includes reflexive and symmetric binary relations.

In 1988, from the approximation retrieval, Lin introduced the notion of the neighborhood systems (NS) as a model of uncertainty in database systems (Lin 1988, 1989), where a neighborhood is a unit of uncertainty. Lin studied the

mathematics foundation of the neighborhood system that originates from the perspective of topology. A topological neighborhood system attaches to every point a collection of subsets that satisfy a set of axioms, called the axioms of topology. Lin removed the axioms and extended the theory to a quite general notion (Lin 1997, 1998, 1999), where a neighborhood system corresponds to a set of binary relationships.

In 1992, Giunchialia and Walsh (1992) presented a theory of abstraction to improve the conceptualization of granularities, where granules are abstracted in hierarchical levels so that relevant elements/attributes can be extracted but irrelevant details are ignored.

In 1997, Zadeh (1997, 1998) refined his information granulation, presented granular mathematics, and applied to word computation. About the same time, Lin used the term “granular computing” to label this growing research field (1996). Since then, granular computing has received more and more attentions, and much research has been conducted in various aspects of this area and has begun to play important roles in various fields, including machine learning and data mining, bio-informatics, e-business, network security, high-performance computing and wireless mobile computing, information hiding, and social works (Bargiela and Pedrycz 2002; Lin 2003; Lin et al. 2002; Pawlak 1998; Skowron and Stepaniuk 2001; Zadeh 1998). International conferences on granular computing sponsored by the IEEE Computational Intelligence Society have been annually held since 2005 (Hu et al. 2005; Zhang and Lin 2006). The essence of these models and applications has been addressed by researchers to build efficient computational algorithms for handling huge amounts of data, information, and knowledge. The objectives of these computation models are computer-centered and mainly concern the efficiency, effectiveness, and robustness of using granules such as classes, clusters, subsets, groups, and intervals in problem-solving. In recent years, some researchers have investigated the granular computing paradigm from perspectives of philosophy, cognitive science, and human thinking as well as the general strategies of interactions

between granules and operations on granule coverings.

Generally speaking, granular computing is a twofold process, granulation and computation, where the former transforms the problem domain to one with granules whereas the latter computes these granules to solve the problem. Granulation is a very natural concept and appears almost everywhere in different names, such as chunking, clustering, partitioning, division, and decomposition, just to name a few. According to Zadeh (1979) and Lin (1998), “information granulation is a collection of granules, with a granule being a clump of objects (points) which are drawn towards an object. In other words, each object is associated a family of clumps.”

Up to now, modeling and applying general GrC is not as successful as desired, although some special models have been innovated with great success. So it is immature at this point to speculate what the general definition, principle, and methodology of granular computing should be. Therefore, we will look at some concrete theories that may be informally called granular computing. The discussion will start from the neighborhood system and extend to the fuzzy set system, the rough set theory, and the quotient space. Then, the perspectives of granular computing in data mining and software engineering will be introduced, including granularity and granulation, granular relationships, and computation with granules.

Perspectives of Granular Computing

Various perspectives of granular computing have been discussed in Yao (2006c, 2016) and Yao et al. (2013), three of which are summarized in Yao (2006c, 2016) from looking at the three levels of holistic and unified view of granular computing.

- From the philosophical perspective, granular computing is a way of structured thinking. It offers general domain independent principles and views for understanding the real world and

problems in the real world using different levels of granularity.

- From the practical perspective, granular computing is a method of structured problem-solving. It provides a general approach to describing, understanding, analyzing, investigating, and solving problems in the real world. With this approach, real-world problems are organized in a hierarchical structure of multiple levels through abstraction, control, complexity, detail, resolution, and so on. With such structures, problems can be examined and resolved using top-down or bottom-up strategies.
- From the computational perspective, granular computing is a paradigm of information processing. It promotes a means of information (including data and knowledge) representation and processing. In this paradigm, granules with internal, external, and contextual properties form larger units, and granular structures provide descriptions of a system or problem under consideration. The information processing consists of granulations and computing with granules, where granulations involve the construction of building granules and structures while computing with granules solves problems by systematically exploring the granular structures. Granular computing as a paradigm of information processing has been recently studied extensively (Pedrycz 2013; Skowron et al. 2016; Yao 2010).

Neighborhood System: A Mathematical Structure of Granular Computing

According to Lin (1997), granular computing can be mathematically structured as a neighborhood system, which is abstracted from the geometric notion of “near” or “negligible distance.” Roughly, a neighborhood system assigns each object in the universe a (possibly empty, finite, or infinite) family of non-empty subsets, called neighborhoods, to represent the semantics of “near.” Neighborhoods play the most fundamental role in granular computing and can be

considered as “granules.” A neighborhood is precisely a topological term of “granules,” and a neighborhood system is vaguely a topological space.

Neighborhood System

Fundamental notions on neighborhood systems were presented by Lin (1997, 1998, 1999). Assume U is the universe of discourse and p is an object in U .

1. An object x is a neighbor of p if x is “near” to p . The notion “near” is a subjective judgment and can be semantically interpreted in different ways, for example, if they satisfy some relationship or condition. For a specific case, let’s consider a binary relationship R in U . x is “near” to p , if xRp .
2. A neighborhood of p , denoted by $N(p)$, in U is a non-empty subset of U , which may or may not contain p . A neighborhood system of p , denoted as $NS(p)$, in U is a family of neighborhoods of p . If p has no neighborhoods, then $NS(p)$ is an empty family.
3. A neighborhood system of U , denoted by $NS(U)$, is the collection of $NS(p)$ for $\forall p \in U$. Formally, $NS(U) = \{NS(p) \mid \forall p \in U\}$.
4. Consider two neighborhood systems of U , say $NS_1(U)$ and $NS_2(U)$. $NS_1(U)$ is said to be a refinement of $NS_2(U)$, if for any neighborhood N_1 in $NS_1(U)$, there exists a neighborhood N_2 in $NS_2(U)$ such that $N_1 \subseteq N_2$. Similarly, $NS_2(U)$ is said to be a weak coarsening of $NS_1(U)$ (see Lin (1998) for strong refinement and coarsening).

From above, one can see that a neighborhood system is induced from the fundamental neighbors of objects, which are imprecisely defined as “near” to. If the “near” is explicitly defined according to a single binary relation, the result neighborhood system is called a binary neighborhood system.

A binary neighborhood system BNS is defined as $BNS = \langle U, B, V \rangle$, where U and V are two universes and B is a mapping function from U to V .

$$B : U \rightarrow V.$$

For $\forall v \in V$, let B_v denote the subset of U that are source of v under the map B , that is,

$$B_v = \{u \mid u \in U \text{ and } B(u) = v\}.$$

Each B_v is called a basic neighborhood or a binary neighborhood.

One can infer that each neighborhood system $BNS = \langle U, B, V \rangle$ defines a binary relation $R \subseteq U \times V, \forall v \in V \text{ and } u \in U, uRv \text{ iff } u \in B_v \subseteq U$. Thus the binary relation R is fully determined by B . Conversely B is fully determined by a binary relation R .

If B is a binary relation $R \subseteq U \times V$, then for each object $v \in V$, the binary neighborhood in terms of v will be

$$N_v = \{u \mid u \in U \text{ and } uRv\}.$$

This is a subset of U , whose all elements are related to v by R .

A binary relation, in the usual sense, is a very special form of a neighborhood system, where each object has either zero or one neighborhood.

If $U = V$ and the mapping function B is considered as a binary relation, then we have a simple binary neighborhood system, denoted as $BNS = \langle U, R \rangle$.

For simplicity, we call this binary neighborhood system with a binary relation as binary relation neighborhood system (BRNS).

Let's consider two binary relation neighborhood systems $BRNS1$ and $BRNS2$ built on the same universes, where $BRNS1 = \langle U, R1, V \rangle$ and $BRNS2 = \langle U, R2, V \rangle$. $BRNS1$ is said to be a weak refinement of $BRNS2$, if $R2 \subseteq R1$, and a strong refinement if every neighborhood of $BNS1$ is a union of neighborhoods of $BNS2$. On the other hand, $BRNS1$ is said to be a weak/strong coarsening of $BRNS2$, if $R1 \subseteq R2$.

Approximation

Assume $BNS = \langle U, B, V \rangle$ is a binary neighborhood system and $X \subseteq U$. X can be approximated by neighborhoods of BNS .

Following topological spaces, the interior approximation, interior for short, of X is defined as either

$$\text{Interior}(X) = \cup_{x \in X} \{N(x) \mid N(x) \subseteq X\}$$

or

$$\text{Interior}(X) = \{x \mid x \in U \text{ and } \exists N(x) \subseteq X\}.$$

Similarly, the exterior approximation, exterior for short, of X can be defined as either

$$\text{Exterior}(X) = \cup_{x \in X} \{N(x) \mid N(x) \cap X \neq \emptyset\}$$

or

$$\text{Exterior}(X) = \{x \mid x \in U \text{ and } \forall N(x) \cap X \neq \emptyset\}.$$

The interior of a subset X of the universe can be approximated by the union of all neighborhoods of the universe that are completely contained in the subset or by a subset of the universe that consists of all objects that have at least one neighborhood contained in X , while the exterior of a subset X can be approximated by the union of all neighborhoods that have non-empty intersection with X or by a subset of the universe that consists of all objects whose all neighborhoods have non-empty intersection with X .

A subset X of U is open if $\forall p \in X$, there is a neighborhood $N(p) \subseteq X$. X is closed if its complement is open. A neighborhood system of p , $NS(p)$, is open if all neighborhoods of p are open. A neighborhood system of U , $NS(U)$, is open, if $\forall p \in U$, $NS(p)$ is open.

If the universe U is a topological space and a neighborhood system of U , $NS(U)$, is open, then $NS(U)$ also defines the same topological space on U . In such a case, both $NS(U)$ and the collection of open sets are called topology.

An object p is a limit point of a set E , if every neighborhood of p contains a point of E other than p . The set of all limit points of E is called derived set. E together with its derived set is a closed set.

One can easily verify that $NS(U)$ is discrete if every $NS(p)$ is a singleton, and $NS(U)$ is indiscrete if $NS(U)$ is a singleton $\{U\}$.

A covering of U is an open neighborhood system $NS(U)$.

A partition of U is a covering of U , where $\forall p, q \in U$, and $p \neq q$, either $NS(p) = NS(q)$ or $NS(p) \cap NS(q) = \emptyset$.

Rough Set System: An Equivalence Neighborhood System

Let's consider a special case of the binary neighborhood system $BNS = \langle U, R \rangle$, where R is an equivalence binary relation on $U \times U$. For an element $u \in U$, by definition, the binary neighborhood of u is defined as

$$N_u = \{x \mid x \in U, \text{ and } xRu\}.$$

Since R is an equivalence relation, N_u is the equivalence class $[u]$ induced by u with respect to R . Thus, the neighborhood system (the collection of all neighborhoods) will be the family of equivalence classes induced by R .

Partition

Rough set theory was proposed by Pawlak in 1982 to deal with inexact information by using rough sets to approximate a crisp set (Pawlak 1982). By investigating the granularity of knowledge from the point of view of the rough set theory, one can view the rough set theory as a granular computing model based on partitions (Pawlak 1998). Basically, suppose U is a finite and non-empty universe. Let $R \subseteq U \times U$ be an equivalence relation on U . The pair $\langle U, R \rangle$ is called an approximation space. The relation R is a special BNS; it can be conveniently represented by a mapping B_R from U to the power set of U and defined as

$$B_R : U \rightarrow 2^U,$$

$$B_R(x) = [x]_R = \{y \in U \mid xRy\}, \forall x \in U.$$

The subset $[x]_R$ is the equivalence class containing x and is called an elementary subset. The family of all equivalent classes, denoted by

$U/R = \{[x]_R \mid x \in U\}$, defines a partition of the universe U , namely, a family of pairwise disjoint subsets whose union covers the whole universe.

From the perspective of the rough set theory, each equivalence class is considered as a whole granule instead of many individuals. With these elementary subsets, any subset of U can be approximated.

Approximation

Let X be a subset of U and R be an equivalence relation on U . The lower approximation of X based on R , denoted by $Low_R(X)$, is defined as

$$Low_R(X) = \cup \{Y \in U/R \mid Y \subseteq X\},$$

which contains all elementary subsets of U that are completely included in X .

The upper approximation of X based on R , denoted by $Upp_R(X)$, is defined as

$$Upp_R(X) = \cup \{Y \in U/R \mid Y \cap X \neq \emptyset\},$$

which contains all elementary subsets of U that have non-empty intersection with X .

In the partition model, these forms of approximations are equivalent to the interior and exterior defined in the BNS, respectively. However, they are not appropriate to be generalized to the NS. For example, in the topological space, $Upp_R(X)$ is always the whole space, no matter what you choose.

Information Systems

The rough set theory has found its applications in data mining, especially in classification and decision-making problem domains, called information systems.

An information system (IS) is defined as: $IS = \langle U, C, D, \{V_{a|a \in C \cup D}\}, f \rangle$, where $U = \{u_1, u_2, \dots, u_n\}$ is a non-empty set of objects (tuples), called data set or decision table, C is a non-empty set of condition attributes, and D is a non-empty set of decision attributes and $C \cap D = \emptyset$. V_a is the domain of attribute a with at least two elements. f is a function: $U \times (C \cup D) \rightarrow V = \cup_{a \in C \cup D} V_a$, which maps each pair of object and attribute to an attribute value.

Let $A \subseteq CUD$ and $t, s \in U$. A binary relation R_A , called an *indiscernibility* relation, is defined as follows: $R_A = \{ \langle t, s \rangle \in U \times U \mid \forall a \in A, t[a] = s[a] \}$, where $t[a]$ indicates the value of attribute $a \in A$ of object t . The *indiscernibility* relation, denoted by $IND(A)$, is an equivalence relation on U . With this equivalence relation R_A , one can construct a binary relation neighborhood system $\langle U, IND(A) \rangle$.

Let X be a subset of U in an information system and represent a concept. X can be lower-approximated and upper-approximated by finding its lower approximation and upper approximation using elementary subsets of $U/IND(A)$. The lower approximation of X contains all objects in U which are *definitely* included in X , while the upper approximation of X contains all objects in U which are *potentially* included in X . On the other hand, one can see that the complement of the upper approximation of X contains all objects in U which are definitely excluded by X . As a concept, X has its lower approximation as the positive region whereas the complement of its upper approximation as its negative region. From the perspective of machine learning or classification, the positive region contains the positive examples of X , while the negative region contains the negative examples of X . Between the positive region and the negative region is called the boundary region of X .

One main application of concept approximation using granules in the rough set theory is to reduce the size of a large decision table by finding an attribute reduct. A condition attribute $a \in C$ is a core attribute of C in U with respect to D if

$$\forall X \in U/IND(D), Low_{IND(C)}(X) \neq Low_{IND(C-\{a\})}(X).$$

A reduct of the condition attribute set is a minimum subset of it that has the same classification capability as itself. Thus a reduct of the condition attribute set can be used to represent the entire condition attribute set.

Formally, a subset A of C , $A \subseteq C$, is defined as a reduct of C in U with respect to D if

$$\begin{aligned} & \forall X \in U/IND(D), Low_{IND(A)}(X) \\ & = Low_{IND(C)}(X), \text{ and} \end{aligned}$$

$$\forall B \subset R, Low_{IND(B)}(X) \neq Low_{IND(C)}(X).$$

A condition attribute $a \in C$ is said to be a reduct attribute if $\exists B \subseteq C$, B is a reduct of C and $a \in B$. A reduct R of C is called a minimum reduct of C if $\forall Q \subset R$, Q is not a reduct of C .

Assume $P \subseteq CUD$ and $Q \subseteq CUD$; the positive region of Q with respect to P , denoted $POS_P(Q)$, is defined as

$$POS_P(Q) = \cup_{X \in U/IND(Q)} Low_{IND(P)}(X),$$

which contains all objects in U that can be classified using the information contained in P . With this definition, the degree of dependency of Q from P , denoted $\gamma_P(Q)$, is defined as $\gamma_P(Q) = POS_P(Q)/|U|$, where $|X|$ denotes the cardinality of the set X .

The degree of attribute dependency provides a measure how an attribute subset is dependent on another attribute subset. $\gamma_P(Q) = 1$ means that Q totally depends on P , $\gamma_P(Q) = 0$ indicates that Q is totally independent from P , while $0 < \gamma_P(Q) < 1$ denotes a partial dependency of Q from P . Particularly, assume $P \subset C$ and $Q = D$; then $\gamma_P(D)$ can be used to measure the dependency of the decision attributes on a subset of conditional attributes. The task of the rough set attribute reduction is to find a subset of the conditional attributes that functions as the original conditional attribute set without loss of classification capability. This subset of the conditional attribute set is called *reduct*. One can prove that $R \subseteq C$ is a reduct of C , if and only if $POS_R(D) = POS_C(D)$, or equivalently, $\gamma_R(D) = \gamma_C(D)$.

Thus, an attribute reduct is such a subset of condition attributes that the decision attribute has the same dependency degree on it as that on the entire set of condition attributes, and no attribute can be eliminated from it without affecting the dependency degree. Given an information system, however, for a given decision table, there may exist more than one reduct. Each reduct can be

used as an alternative group of attributes to represent the original information system. It has been proved that every reduct must contain all the core attributes. However, finding all reducts of the condition attribute set is unfortunately NP-hard.

Quotient Space

Multiple reducts of condition attributes in an information system may exist. Each reduct can be used to define an equivalence relation. Thus it can be applied to construct a quotient space. From the perspectives of mathematics and topology, a quotient space is a space where equivalent points are glued together, and therefore a new space can be constructed.

Assume the universe of discourse U is a topological space; R is an equivalence relation on U . U/R , the set of all equivalence classes of R , is called a quotient space, where the topology on U/R is defined as below:

A subset $X \subseteq U/R$

is open if and only if their union is open.

Topological properties of a quotient space as well as relationships between quotient spaces have been studied to identify and select attributes to form high-quality reducts. The quotient space theory has also been explored to solve information hiding, network security, and other problems (Zhang and Zhang 2003).

Fuzzy Set System: A Fuzzy Neighborhood System

In this section, another special case of binary neighborhood system is considered. Assume a binary neighborhood system $BNS = \langle U, B, V \rangle$, where U is a universe of discourse (the data space), V is a concept space, and B is a mapping function from $U \times V$ to the unit interval. Formally,

$$B : U \times V \rightarrow [0, 1].$$

Since each point in the concept space is actually a concept, and can be assigned a name, Lin

(1999) suggested that this neighborhood system be defined as a four-tuple system $\langle U, B, V, C \rangle$, where U , B , and V are the same as before while C is the concept name space. This system is called a fuzzy granular structure, and each concept in the concept space is a fundamental fuzzy clump or fuzzy neighborhood. In the simple case, B is a fuzzy binary relation, and thus the four-tuple system above is a single-level granulation, called a fuzzy binary granular structure.

Information Granularity

Information granularity with fuzzy set was proposed by L. A. Zadeh in 1979 (Zadeh 1979) to provide a basis for the construction of more general theories in which the evidence is allowed to be fuzzy in nature. In the information granularity, the data granules are viewed as a proposition in the general form of

$$g : X \text{ is } G \text{ is } \lambda,$$

where X is a variable taking values in a universe of discourse U , G is a fuzzy set of the universe and is characterized by its membership function μ_G , and λ is a fuzzy probability characterized by a possibility distribution over a unit interval.

A typical example of such a proposition is given in the universe U of real numbers, as below:

$$g : X \text{ is small is likely},$$

where the fuzzy subset G is “small” and the fuzzy probability λ is a fuzzy subset of the unit interval.

Each fuzzy subset G of the universe can be generally understood as a concept, for example, “small,” “very large,” “old,” “young,” etc. All concepts are put together to form a concept space, denoted by C

$$C = \{G_1, G_2, \dots, G_N\}.$$

Each proposition g as above is considered as an evidence, and all evidences comprise a collection of propositions:

$V = \{g_1, g_2, \dots, g_N\}$, where $g_i : X \text{ is } G_i \text{ is } \lambda_i$,
 $i = 1, 2, \dots, N$.

The proceeding information granularity can be characterized as a neighborhood system. Consider a mapping function B from $U \times V$ to $[0, 1]$, defined as

$$B : U \times V \rightarrow [0, 1], B(x, g) \\ = \mu_{\lambda}(\mu_{G(g)}(x)), \forall x \in U \text{ and } \forall v \in V.$$

where $\mu_{G(g)}$ is the membership function of the concept G in the proposition g . Thus the four-tuple $\langle U, B, V, C \rangle$ forms a neighborhood system in the form of fuzzy binary granular structure.

A special case of fuzzy granules is called non-probability-qualified granule and in the form of “ X is G .” In this case, the corresponding fuzzy neighborhood system can be rewritten as a four-tuple system $FNS = \langle U, B, V, C \rangle$, where U is the universe of discourse (data space), V is a family of fuzzy subsets of U , C is the name space whose names correspond to the fuzzy subsets in V , and B is a mapping function:

$$B : U \times V \rightarrow [0, 1], B(x, v) \\ = \mu_v(x), \forall x \in U \text{ and } \forall v \in V.$$

Consider the following example: $U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $V = \{V_1, V_2, V_3, V_4\}$, where (using discrete fuzzy subset notation)

$$V_1 = 1.0/0 + 0.67/1 + 0.33/2 + 0/3 + 0/4 \\ + 0/5 + 0/6 + 0/7 + 0/8 + 0/9,$$

$$V_2 = 0/0 + 0.33/1 + 0.67/2 + 1.0/3 + 0.67/4 \\ + 0.33/5 + 0/6 + 0/7 + 0/8 + 0/9,$$

$$V_3 = 0/0 + 0/1 + 0/2 + 0/3 + 0.33/4 \\ + 0.67/5 + 1.0/6 + 0.67/7 + 0.33/8 \\ + 0/9,$$

$$V_4 = 0/0 + 0/1 + 0/2 + 0/3 + 0/4 + 0/5 \\ + 0/6 + 0.33/7 + 0.67/8 + 1.0/9,$$

and the collection of words $C = \{\text{Low, Median-Low, Median-High, High}\}$ to name fuzzy subsets V_1, V_2, V_3 , and V_4 .

Conditional granules can be represented as fuzzy rules such as “If $X = u$, then Y is G .” These kinds of granules can also be represented as fuzzy set membership functions or expressions of these functions (Lin 1999).

Approximation

In a fuzzy neighborhood system $FBS = \langle U, B, V, C \rangle$, the concept space V (a family of fuzzy subsets of U) is also called a fuzzy covering of U , and each v in V is called a cover of U . For each $u \in U$, $v \in V$ is called a fundamental fuzzy neighborhood of u , if $B(u, v) = \mu_v(u) > 0$. Since each object u in the data space U may have multiple fundamental fuzzy neighborhoods, u can be represented by a vector of these fundamental fuzzy neighborhoods, called multivalued representation for multilevel granulation. All fundamental fuzzy neighborhoods of u put together is denoted by $FNS(u)$.

Given a fuzzy neighborhood system $FBS = \langle U, B, V, C \rangle$, a subset X of U can be approximated by a membership function. The minimum approximation $\min(X)$ of X in FBS can be defined as the following membership function:

$$\mu_X(x) = \min \{\mu_v(x) | v \in FNS(x)\} \\ \text{if } x \in X, \text{ and } 0 \text{ otherwise.}$$

Similarly, the maximum approximation $\max(X)$ of X in FBS can be defined by the following membership function:

$$\mu_X(x) = \max \{\mu_v(x) | v \in FNS(x)\} \\ \text{if } x \in X, \text{ and } 0 \text{ otherwise.}$$

One can see from above that the minimum (maximum) approximation of a subset of U is a fuzzy subset of U that minimizes (maximizes) the membership degrees of its elements to their fuzzy fundamental neighborhoods.

Computing with words is to make inferences from words to words and is a very important

topic of information granularity (Zadeh 1998). Fundamental neighborhoods can also be linearly combined to approximate formal words (concept names) (Lin 1999). Basically, let $C = \{C_1, C_2, \dots, C_m\}$ be the concept space of a fuzzy neighborhood system and $V = \{V_1, V_2, \dots, V_m\}$ be the corresponding fuzzy subsets of the universe U . Assume $r_1, r_2, \dots, r_m$ are real numbers. Consider the following linear expression:

$$r_1 C_1 + r_2 C_2 + \dots + r_m C_m = \sum_{i=1, \dots, m} r_i C_i.$$

Mathematically, the collection of all such expressions forms an abstract vector space, and each vector in the space is called a formal word, corresponding to a new concept. This new concept can be approximated by a fuzzy subset of U , of which the membership function is defined in terms of the membership functions μ_{V_i} , $i = 1, 2, \dots, m$, for example, the minimum approximation or the maximum approximation discussed previously. The membership function of the new concept can also be defined as a linear combination of membership functions of its all fundamental fuzzy neighborhoods, i.e.,

$$\mu_{\text{New Concept}}(x) = \sum_{i=1, \dots, m} r_i \mu_{V_i}(x).$$

In the above example, a concept Median can be defined in the middle between the concepts Median-Low and Median-High, that is,

$$\begin{aligned} \text{Median} = & 0 * \text{Low} + 1/2 * \text{Median} - \text{Low} \\ & + 1/2 * \text{Median} - \text{High} + 0 * \text{High}. \end{aligned}$$

This concept (word) can be approximated by the fuzzy subset defined below:

$$\mu_{\text{Median}} = 1/2 * \mu_{V_2} + 1/2 * \mu_{V_3},$$

and the fuzzy subset is

$$\begin{aligned} & 0/0 + 0.16/1 + 0.33/2 + 0.5/3 + 0.5/4 \\ & + 0.5/5 + 0.5/6 + 0.33/7 + 0.16/8 + 0/9. \end{aligned}$$

Granular Computing in Data Mining

Granular computing, as a strategy of problem-solving and human thinking, has been extensively applied in knowledge discovery and data mining. For example, machine learning is to interpolate data and integrate concepts based on background knowledge that is viewed as granules. In neural network computing, activation functions can also be regarded as granules, and thus the propagation and processing of activation functions can be viewed as a form of computations with granules. A gene is a granule, genetic algorithms process and transform genes which are actually granules, and therefore genetic algorithms can also be considered as one form of granular computing. In this section classification and clustering analysis and especially association mining are reviewed from the perspective of granular computing.

Granular Computing in Classification and Clustering

Classification problem is, given a training set of examples with class labels, to construct a classifier that is able to assign a class label to a new example without class label that is not in the training set (Gordon 1981). All examples are described with a given set of attributes. On the other hand, clustering is partitioning a data set by the similarity into clusters. Classes in classification and clusters in clustering are granules, subsets of examples, and the granulation is to partition the set of examples into granules. Computation with classes is to induce classification rules to characterize each class and predict the classes of future examples (James 1985). Similarly, computation with clusters is to summarize or characterize each cluster and discriminate between clusters.

The classification can be based on the refinement and coarsening relationships between granules (Yao 2006a). A granule g_1 is defined as a refinement of another granule g_2 , or equivalently, g_2 is a coarsening of g_1 , if every subgranule or object of g_1 is contained in some subgranules of g_2 . Partitions and coverings are two simple and commonly used granulations of a universe. A partitioning of a universe is a collection of its

non-empty and pairwise disjoint subsets, whose union is the universe. On the other hand, a covering of a universe is a collection of its non-empty subsets that cover the whole universe. A partitioning is a special case of covering.

On the other hand, the clustering can be built upon the similarity relationship (Yao 2006a). Similarity between granules is a basic and fundamental granular relationship and a key to forming an intrarelation of a granule. Furthermore, it can be used to measure closeness or nearness among granules. Various measures can be exploited to calculate the similarity between two granules, which can be defined as the average distance between individuals in the two granules, and the distance between individuals can be measured as, for example, the Euclidean distance between points, the Levenshtein edit distance between strings or texts, the amino acid distance in biology, and the Mahalanobis squared distance between statistics data items.

Granular Computing in Association Mining

Discovering association rules is another important area of data mining. The basic task of mining association rules is to find frequent itemsets which are granules and then extract associations or correlations between items in the frequent itemsets, which is the computation with granules.

An association rule is an implication of the form $A \rightarrow B$, where A and B are subsets of regarded attributes, and $A \cap B = \emptyset$. For different types of attribute values, the implication of the association form may have different meanings. The problem of mining association rules can be interpreted from the perspective of granular computing.

Mining association rules in large databases were first presented by Agrawal, Imielinski, and Swami in 1993 (Agrawal et al. 1993) to solve the basket data problem, in which, given a large database of customer transactions (basket data) that consist of items purchased by customers, significant associations between items are pursued. For example, “most often transactions that purchase bread and butter also purchase milk,” and “a customer purchasing tea is likely to also purchase coffee.” These significant associations are

described as a set of association rules and quantitatively measured with support and confidence.

Formally, let $I = \{I_1, I_2, \dots, I_m\}$ be a set of items, each I_k being an item, where m is the number of items. Let D be a database of transactions, where each transaction corresponds to a tuple in D , denoted by T . A transaction T is a set of items represented as a binary vector, with $T[k] = 1$ if T contains the item I_k , and $T[k] = 0$ otherwise, for $1 \leq k \leq m$. Assume $X \subseteq I$ is a subset of items in I . A transaction T satisfies X if for all items $I_k \in X$, $T[k] = 1$.

In the basket data problem, an association rule is of the form $X \rightarrow Y$, where $X, Y \subseteq I$, and $X \cap Y = \emptyset$. The rule $X \rightarrow Y$ holds in D with confidence c if at least $c\%$ transactions in D that contain X also contain Y and support s if at least $s\%$ transactions in D contain $X \cup Y$.

The task of mining association rules in D is to generate above implications with support and confidence greater than or equal to the user-specified support threshold δ and confidence threshold σ , respectively. Confidence is a measure of the rule strength, while support corresponds to statistic significance.

The problem of mining association rules can be decomposed into two subproblems:

1. Discovering all frequent itemsets: a subset X of I is a frequent itemset if its support is at least δ .
2. Generating association rules: for each frequent itemset X , construct implications of the form $X - Y \rightarrow Y$, for all $Y \subset X$. If the confidence of such a form is at least σ , it is considered as an association rule.

Currently many approaches to solving the basket data problem have been developed, and efficient algorithms have been implemented. The classical solution exploits level-wise search to generate all candidate frequent itemsets and then prunes them (Agrawal and Srikant 1994). The level-wise methods are based on the observation that if an itemset is frequent, then all of its subsets are frequent. Thus, the basic idea is to evaluate 1-itemsets first, find frequent singleton itemsets, and then evaluate their supersets 2-itemsets and so forth. Such evaluations are repeated until no

frequent itemset is found or the size of the itemsets approaches some threshold.

Louie and Lin reformulate the association relations into bit data model (Louie and Lin 2000), and Lin also reformulates these relations into granular model (Lin 2000) which has been extended to a theory of attributes (features) (Lin 2002).

From the perspective of granular computing, association rule mining can be formalized as a binary neighborhood system $BNS = \langle I, B, D \rangle$, where the universe of discourse is the set I of all items; transaction database D (the collection D of tuples) is a subset of the power set of I , that is, $D \subset 2^I$; and the binary relation B can be understood as, for all items i and j , iBj , if and only if there exists a tuple in D that contains both i and j . D forms a covering of I , but not a partition of I .

An itemset $X \subset I$ can be approximated by a subset $C(X)$ of D , where C is defined as

$$C(X) = \{Y \mid Y \in D, X \subseteq Y\}, \text{ for all } X \subseteq I.$$

X is a frequent itemset, if the cardinality of $C(X)$, denoted by $|C(X)|$, is greater than or equal to the support threshold δ . A frequent itemset is called granule. The level-wise search process can be applied to find all granules.

The computation with granules in generating association rules is mainly within granules. For each subset Y of such a granule (a frequent itemset) X , one can find the approximation $C(X)$ and $C(Y)$ of X and Y , respectively. The confidence of implication $X \rightarrow Y$ can be calculated as $|C(X)|/|C(Y)|$.

A Framework for Granular Computing in Data Mining

Yao (2006b, 2008) has been studying various aspects of granular computing for data mining and proposed a framework for granular computing in data mining. His proposal is based on the following assumptions, which are related to the basic components and task of granular computing.

- Knowledge granule: each granule represents a piece of knowledge, corresponding to granules.

- Structural knowledge: connection between a web of knowledge granules represents structural knowledge, corresponding to the granular structure.
- Mining task: searching for meaningful knowledge granules and structural knowledge is a basic task of data mining, corresponding to granulation and computing with granules.

Yao (2006b) relates granules to the well-studied notion of concepts and adopts the classical view of concepts, in which every concept is understood as a unit of thought that consists of two parts, the intension and the extension of the concepts. A language can be defined so that the intension of a concept is expressed as a formula of the language and the extension is the set of objects satisfying the formula. This way, concepts can be studied in a logic setting in terms of intensions and also in a set-theoretic setting of extensions. In data mining and machine learning, extensions of concepts are normally defined with respect to a particular training set of examples, while intensions of concepts can be represented as (various kinds of) rules.

Yao (2006b) also pointed out that knowledge granules serve as building blocks to derive structural knowledge, and human knowledge can be characterized using context and hierarchy, such as formal concept lattices (Ganter and Wille 1999) and hierarchical classes (De Boeck and Van Mechelen 1990). Therefore, structural knowledge can be expressed in terms of connections between a web of knowledge granules.

Granular Computing in Software Engineering

In software engineering, the strategies of granular computing have been broadly used in all phases. The granulate-and-conquer is a softer version of classical divide-and-conquer strategy that employs recursive algorithms to solve problems. A very common technique used in the classical “non-partitioning” recursive call is dynamic programming. Functional decomposition is to

partition user requirement into granules (functions). Structured programming is to organize the computer programs as a collection of modules (procedures, functions, routines). The intra-relationships among these granules are based on their ingredients such as input parameters, output results, and executable statements, whereas the interrelationships involve the module interfaces and procedure calls. In object-oriented programming, granules are classes; intra-relationships are the interactions between components of classes such as instance fields, methods, and constructors in the Java language; and the interrelationships are class inheritance, aggregation, association, delegation, dependency, etc.

In this section, the different phases of modern software engineering process are investigated from the granular computing perspective, based on object-oriented methodology (O'Docherty 2005), including requirement analysis, system analysis, design, and implementation.

User Requirements

In the object-oriented methodology, user requirement analysis involves two main aspects: identifying business actors and use cases. Both aspects can be conducted with granular computing paradigm. However, only identifying use cases will be investigated in this subsection.

Basically, each use case is a snippet of the business and may involve a two-way communication between actors and the system. If the business requirement is considered as the problem domain, then the use cases will be the granules. Although there's no set of rules for deciding how to granulate the business into use cases, most analysts partition the business process into use cases based on common sense, business logic, and their experience. The basic ingredients of the business requirement analysis from the granular computing perspective can be summarized as follows.

Granules: use cases.

Granulation method: top-down decomposition for complete covering and bottom-up combination to form a hierarchy of use cases.

Granular relationships: some relationships between use cases can be recognized. For

example, inheritance is a specialization and generalization relationship in which a large case is decomposed into small cases or a set of small cases are combined to constitute a large case. Inheritance is often referred to as an is-a relationship: a granule g_1 is a special case of another granule g_2 . The is-a relationship plays a very important role in the object-oriented methodology and concept analysis, since " g_1 is-a g_2 " reveals that g_1 inherits all features and functions of g_2 (Yao 2006a). Other relationships between use cases include inclusion where the source case has some of its steps provided by the target case and extension where the source case adds steps to the target case (O'Docherty 2005).

Granular computation: identifying the process of use cases, including input data, output data, and business logic, as well as the communication with actors.

System Analysis

The basic goal of system analysis is to find candidate classes that describe the objects that might be relevant to the system, relationships between the classes, as well as attributes for the classes. With granular computing terminology, classes are granules that all together should cover the system requirement and satisfy the needs of use cases that have been identified before. The ingredients of system analysis from the granular computing perspective are summarized as follows:

Granules: classes.

Granulation method is to identify classes. Candidate classes are often indicated by nouns in the use cases except those that represent the system, actors, boundaries, and trivial types. Two basic rules should be followed to identify classes: high cohesion inside classes and low coupling between classes. Class identification might be top-down and/or bottom-up, but the general complete coverage is necessary.

Granular relationships correspond to class relationships, which can include the following types (O'Docherty 2005): inheritance (is-a relationship) where a subclass inherits all of the attributes and behavior of its superclass(es); association, where objects of one class are associated with objects of another class; aggregation

(strong association), where an instance of one class is made up of instances of another class; composition (strong aggregation) where the composed object can't be shared by other objects and dies with its composer; and others as well.

Computation with granules is twofold. First, determine the internal structures and behaviors of objects of classes, including the instance fields and method/constructor prototypes; and second, interface the connections between classes. Associating classes often affects the design of internal structures of classes.

System Design

Basically the software system design is to decompose a system into physical and logical components and determine the technologies to be used to implement the system. Traditional system design focuses on the system functional partitioning, while modern software system design is based on object-oriented technology. In this subsection, object-oriented system architecture design and technology are discussed from the perspective of granular computing.

System design involves multiple steps, including system topology, software partitioning, concurrency and security policies, as well as communications between components. The following discussion will concentrate in the topology design and system partitioning.

The first task of current object-oriented system design is normally to determine the topology of a networked system. One popular system topology is the three-tier or multi-tier architecture to separate user interfaces, program logic, and data in the system. In the three-tier architecture, the client tier presents the user interface to the user so that he/she can enter data and view results; the middle tier – also known as the business logic tier or server tier – runs a multi-thread program code using large processors and lots of memory; and the data tier stores the data and provides safe concurrent access to it, typically with the help of a database management system. Thus, the system is granulated into three components. Besides the design of these three tiers, the protocols between tiers are also important.

To granulate the software, thus, object-oriented system design partitions the system components into layers: *user interface*, *control*, *network*, *server*, *business*, *persistence*, and *database*.

System Implementation

Software system implementation can also be viewed from the perspective of granular computing. The typical program structure in object-oriented software system is actually a granular structure. Consider a Java-based software system, which can be characterized as follows:

Software system
Packages
Classes
Instance fields
Constructors
Methods
Interfaces
Method prototypes

In this structure, a software system is designed as a set of packages, which may be installed in different physical devices. Each package contains a set of interfaces whose subgranules are method prototypes and a set of classes which is partitioned into three types of components: instance fields, constructors, and methods.

During the design of classes, algorithms are one of designer's main concerns. Granular computing is also an important strategy for this purpose. Let's consider the divide-and-conquer technique that is frequently used in algorithm design. The general paradigm of divide-and-conquer (Goodrich and Tamassia 2004) is

Divide: divide the problem S in two or more disjoint subsets $S_1, S_2, \dots$

Recur: solve the subproblems recursively.

Conquer: combine the solutions to $S_1, S_2, \dots$, into a solution to S .

The base cases for the recursion are subproblems of constant size.

In the above paradigm, each subproblem is a granule, which can be granulated further into smaller subproblems. Granulation process usually follows the top-down method, and the obtained granules should completely cover the parent problem. The granularity constitutes a hierarchy,

where the top is the problem originally given, while at the bottom are all base cases. The computing with granules consists of not only dividing a non-basis problem or solving a base problem but also combining solutions to subproblems to form a solution to the parent problem.

Another important general algorithm design paradigm is dynamic programming, which is actually a special case of divide-and-conquer and thus a granular computing strategy. The main difference between them is: divide-and-conquer solves each subproblems individually, while dynamic programming stores solutions to all subproblems so that when a subproblem is reencountered, the solution can be obtained directly without resolving it.

Future Directions

In this entry, the basic principles and models of granular computing were briefly summarized. The ingredients of granular computing were investigated in terms of granulation methods and criteria, granular relationships, and computations with granules as well. Data mining technologies, especially association and correlation mining, and object-oriented technologies were reviewed from the perspective of granular computing. Granular computing methodology has attracted researchers and practitioners from various fields, but the unified architecture and generally accepted framework of granular computing have not been established. It is still an open question what granular computing is and how granular computing is performed. To answer such kind of questions, the future research on granular computing will be focusing on the following directions.

Foundations of granular computing: The mathematical foundation of granular computing will be one of the important research topics, which will develop the granular structures of granules, granular relationships between granules, and computations with granules from the point of view of mathematics. Neighborhood systems, fuzzy set theory, rough set theory, and quotient space theory have initialized this research direction, but from different concerns and computing

tasks. Granulation is based on partitioning and covering in terms of the nature of applications. Although it is difficult to propose a general process for granulation and in most applications granulation is performed by human experts, the basic principles and rules will be built for practitioners to follow. For example, granulation should meet the universal approximation properties. The infinite input space is granulated into finite granules, but finite granules should be able to approximate the universe of discourse in some ways.

Framework of granular computing: The hierarchical structure of granular computing will be established to represent the granules, relationships between granules, and computation with granules from the point of view of machine-centered computations. This research is originated from human thinking and problem-solving and will be developed from philosophy, psychology, and behavior sciences. The general framework of granular computing will concentrate in the granulation and granule transformation.

Information integration: Basically, granular computing granulates complex problems into subproblems, conquers subproblems, and integrates solutions to subproblems to obtain the solution to the original complex problem. Granulate-and-conquer is the natural extension of the divide-and-conquer strategy and addresses the integration of the subsolutions on granules into the total solutions on the original problem. For general granular computing, it is possible that two distinct problems be granulated into the same collection of granules. Different integrations of the same subsolutions may lead to the different results.

Applications of granular computing: Granular computing is strongly domain dependent. For example, in fuzzy logic control, the unknown control functions can be expressed by the membership functions of fuzzy granules; in structural programming, the programmers partition the functions according to their experience and background knowledge, while in rough set theory, the indiscernibility equivalence relation is provided a priori by experts and unknown concepts are approximated by equivalence classes. These examples demonstrate that granulation is not algorithmic and strongly domain dependent.

Similarly, computation with granules is dependent on the domain knowledge and the computation task. Extracting the common properties and features of granular computing in various application domains will be emphasized to enhance the general framework of granular computing and contribute to the foundation of granular computing.

Bibliography

Primary Literature

- Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20th international conference on very large data bases, Santiago, pp 487–499
- Agrawal R, Imielinski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD conference, Washington, DC
- Bargiela A, Pedrycz W (2002) Granular computing: an introduction. Kluwer Academic, Boston
- De Boeck P, Van Mechelen I (1990) Traits and taxonomies: a hierarchical classes approach. *Eur J Personal* 4:147–156
- Ganter B, Wille R (1999) Formal concept analysis: mathematical foundations. Springer, New York
- Giunchiglia F, Walsh T (1992) A theory of abstraction. *Artif Intell* 56:323–390
- Goodrich M, Tamassia R (2004) Algorithm design: foundations, analysis, and internet examples, 2nd edn. Wiley, New Jersey, USA
- Gordon AD (1981) Classification. Chapman Hall, London
- Hobbs J (1985) Granularity. In: Proceedings of international joint conference on artificial intelligence, pp 432–435
- Hu X, Liu Q, Skowron A, Lin TY, Yager RR, Zhang B (2005) Proceedings of IEEE international conference on granular computing. IEEE Press, Los Alamitos
- James M (1985) Classification algorithms. William Collins, Glasgow
- Lin TY (1988) Neighborhood systems and relational database. Abstract. In: Proceedings of CSC'88, p 725
- Lin TY (1989) Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings of the fourth international symposium on methodologies of intelligent systems, pp 75–86
- Lin TY (1997) Neighborhood systems – a qualitative theory for fuzzy and rough sets. In: Wang P (ed) Advances in machine intelligence and soft computing, vol IV. Duke University, North Carolina, pp 132–155
- Lin TY (1998) Granular computing on binary relations I: data mining and neighborhood systems. In: Skowron A, Polkowski L (eds) Rough sets in knowledge discovery. Physica-Verlag, New York, pp 107–121
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh LA, Kacprzyk J (eds) Computing with words in information/intelligent systems. Physica-Verlag, New York, pp 183–200
- Lin TY (2000) Data mining and machine oriented modeling: a granular computing approach. *Appl Intell* 13(2):113–124
- Lin TY (2002) Attribute (feature) completion – the theory of attributes from data mining prospect. In: Proceedings of international conference on data mining, pp 282–289
- Lin TY (2003) Granular computing. In: Lecture notes in computer science, vol 2639. Springer, Berlin, pp 16–24
- Lin TY, Yao YY, Zadeh LA (eds) (2002) Data mining, rough sets and granular computing. Physica-Verlag, Heidelberg
- Louie E, Lin TY (2000) Finding association rules using fast bit computation: machine-oriented modeling. In: Proceedings of international symposium on methodologies for information systems, pp 486–494
- O'Docherty M (2005) Object-oriented analysis and design. Wiley, New Jersey, USA
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11:341–356
- Pawlak Z (1998) Granularity of knowledge, indiscernibility and rough sets. In: Proceedings of IEEE international conference on fuzzy systems, pp 106–110
- Pedrycz W (2013) Granular computing analysis and design of intelligent systems. CRC Press, Boca Raton
- Skowron A, Stepaniuk J (2001) Information granules: towards foundations of granular computing. *Int J Intell Syst* 16:57–85
- Skowron A, Jankowski A, Dutta S (2016) Interactive granular computing. *Granul Comput* 1(2):95–113
- Yao J (2006a) Information granulation and granular relationships. In: Proceedings of IEEE GrC, pp 326–329
- Yao Y (2006b) Granular computing for data mining. In: Proceedings of SPIE conference on data mining, intrusion detection, information assurance, and data networks security, La Jola
- Yao Y (2006c) Three perspectives of granular computing. *J Nanchang Inst Technol* 25(2):16–21
- Yao Y (2008) A unified framework of granular computing. In: Pedrycz W, Skowron A, Kreinovich V (eds) Handbook of granular computing. Wiley, Hoboken, pp 401–410
- Yao Y (2010) Human-inspired granular computing. In: Yao J (ed) Novel developments in granular computing: applications for advanced human reasoning and soft computation. IGI Global/Hershey, Philadelphia, pp 1–15
- Yao Y (2016) A triarchic theory of granular computing. *Granul Comput* 1(2):145–157
- Yao J, Vasilakos AV, Pedrycz W (2013) Granular computing: perspectives and challenges. *IEEE Trans Cybern* 43:1977–1989

- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yager R (eds) *Advances in fuzzy set theory and applications*. North-Holland, Amsterdam, pp 3–18
- Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 19:111–127
- Zadeh LA (1998) Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems. *Soft Comput* 2:23–25
- Zhang Y, Lin TY (2006) *Proceedings of IEEE international conference on granular computing*. IEEE Press, Los Alamitos
- Zhang L, Zhang B (2003) The quotient space theory of problem solving. In: *Lecture notes in computer science*, vol 2639. Springer, Berlin, pp 11–15

Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities

De Wang¹, Steve Webb¹, Kyumin Lee², James Caverlee³ and Calton Pu¹

¹College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

²Department of Computer Science, Utah State University, Logan, UT, USA

³Department of Computer Science, Texas A&M University, College Station, TX, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Background on Social Networking Communities

Traditional Attacks Targeting Social Networking Communities

Malware Propagation

Spam

Phishing

New Attacks Against Social Networking Communities

Social Spam

Campaigns

Misinformation

Future Directions

Bibliography

Glossary

Social networking community A community of individuals (users) that are connected based on mutually shared activities, beliefs, goals, or relationships.

Profile A user-controlled web page that contains a picture of a community user along with various pieces of personal information for that

user. This is often the online representation of an individual in a social networking community.

Friend request A communication mechanism that allows community users to initiate connections between their profiles and the profiles of their friends.

Following A communication mechanism that allows community users to subscribe updates (e.g., messages, images, videos) and receive direct information from other users.

Definition of the Subject

Online social networking communities are connecting hundreds of millions of individuals across the globe and facilitating new modes of interaction. Due to their immense popularity, an important question is whether these communities are safe for their users. In this entry, we address this safety question and show that social networking communities are susceptible to numerous attacks. Specifically, we identify two attack classes: traditional attacks that have been adapted to these communities (e.g., malware propagation, spam, and phishing) and new attacks that have emerged through malicious social networking profiles (e.g., social spam, campaign, and misinformation). Concretely, we describe examples of these attack types that are observable in Facebook, Twitter, and Myspace, which are or used to be popular social networking communities.

Introduction

Over the past few years, social networking communities have experienced unprecedented growth. Communities such as Facebook, Twitter, and Myspace are connecting people in a variety of new and exciting ways, and as a result, individuals are attaching an increasing amount of value to their online personas. Unfortunately, the rising

importance and prominence of these communities have also made them prime targets for attack by malicious entities.

We observe two distinct attack classes that threaten social networking communities and the privacy of their users. First, traditional attacks that have plagued Internet users for many years (e.g., malware propagation, spam, and phishing) have been adapted to take advantage of the unique properties of online communities. These traditional attacks are thriving in their new environment, and the severity of these attacks promises to grow as attackers become more sophisticated. The second attack class consists of new attacks that have emerged from the very fabric of these communities.

In this entry, we provide detailed descriptions for each of these attack classes, and we show that the continued success of social networking communities is contingent upon their ability to mitigate the risks associated with these attacks. Our observations show the practical importance of these attacks and their impact on millions of users. Additionally, since other social networking communities are both functionally and structurally similar to Facebook, Twitter, or MySpace, the attacks we describe can be easily adapted to most (if not all) social networking communities.

The rest of the entry is organized as follows. Section “[Background on Social Networking Communities](#)” provides background information about social networking communities. In section “[Traditional Attacks Targeting Social Networking Communities](#),” we describe traditional attacks that have adapted to target social networking environments, including malware propagation, spam, and phishing. In section “[New Attacks Against Social Networking Communities](#),” we present new attacks that specifically target social networking communities. Section “[Future Directions](#)” concludes the entry.

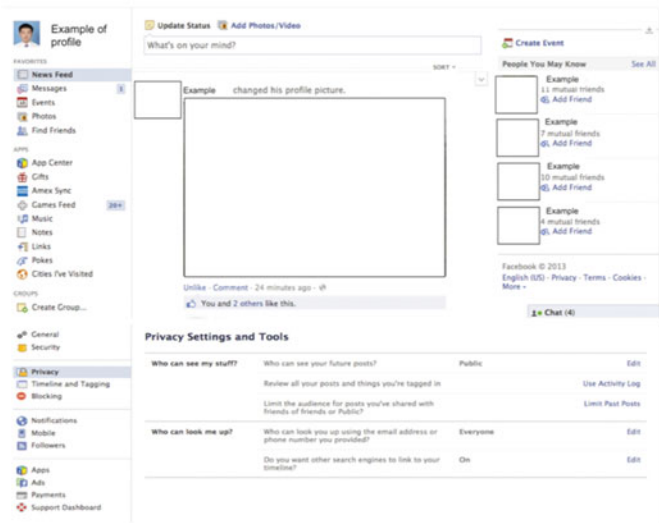
Background on Social Networking Communities

Social networking communities provide an online platform for people to manage existing

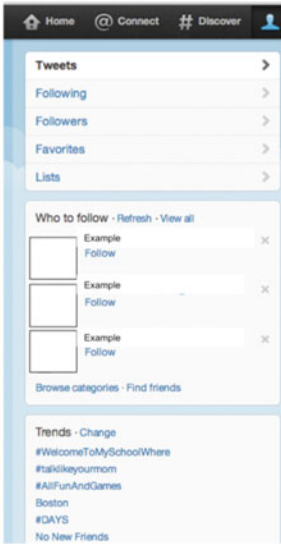
relationships, form new ones, and engage in a variety of social interactions. Typically, a user’s online presence in these communities is represented by profile, which is a user-controlled web page that contains a picture of the user along with various pieces of personal information. Profiles connect to other profiles through explicitly declared friend relationships and numerous messaging mechanisms.

Examples of Facebook and Twitter profile are shown in Fig. 1. Some of a profile’s personal information is mandatory (e.g., the user’s name, age, gender, location), and some of it is optional (e.g., the user’s interests, relationship status, occupation). In addition to personal information and embedded content, a user’s profile also contains a list of various Apps to use and groups to create or join. Meanwhile, users could create or join events through event links and find friends through the recommendation friend links. For these friend links, they are bidirectional because a link is established only after both parties acknowledge the friendship. To initiate a friendship, a user sends a friend request to another user. If the other user accepts this request, the friendship is established, and a friend link is added to both users’ profiles. For purpose of security and privacy, a user could set up privacy policies and use tools provided by Facebook. For instance, who can see my stuff? who can look me up? While Twitter has different communication mechanism in which the friend links are replaced with following links. Users can follow anyone on Twitter without the acknowledgment of the targets. Also, Twitter profile contains a list of followers who subscribe this users’ tweets and trending topics that are the latest top 10 popular topics.

Aside from friend requests, social networking communities provide a number of other communication facilities that enable users to communicate with each other within the community. These facilities include messaging, bulletin, commenting, blogging, and instant messaging (IM) systems. The messaging system allows users to exchange intracommunity email messages with any other user (i.e., both friends and strangers). The bulletin system is essentially an exclusive bulletin board that only a user’s friends can view, enabling users



(a) Facebook profile and privacy settings



(b) Twitter profile

Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities, Fig. 1 Example Facebook and Twitter profiles. In

(a), the profile is public profile example in Facebook. It also shows the example privacy setting from Facebook. In (b), the profile is public profile example in Twitter

to communicate with all of their friends at once. Users can post comments on their friends' profiles using the commenting system, and the blogging system allows users to maintain blogs on their profiles, which other users can read and comment on. Finally, the IM system provides users with a mechanism to send intracommunity instant messages to any other user.

Due to the wealth of private information that is accessible on user profiles and the various means of communication that are available, Facebook or other similar online social network communities provide mechanisms to protect the privacy of its users. First and foremost, users have the ability to choose between making their profiles publicly viewable (the default option) or private. If a user's profile is designated as private, only the user's friends are allowed to view the profile's detailed personal information (e.g., the user's interests, blog entries, comments). They also provide a few finer-grained privacy mechanisms. Users can control who is allowed to IM them (everyone, only friends, or no one) and who is allowed to leave blog comments (everyone or

only friends). Users can also maintain a block list to prevent specific users from contacting them at all. Unfortunately, as we will see in the following sections, these privacy mechanisms have been unable to prevent a number of attacks.

Traditional Attacks Targeting Social Networking Communities

Since social networking communities include communication facilities that are fundamentally similar to traditional means (i.e., email, blogs, instant messaging), many of the attacks that are effective against those traditional communication media have been adapted to exploit social network communications. Due to the massive size of many social networking communities, their tightly connected nature, and their relatively naïve user bases, these communities are target-rich environments for attackers. In this section, we describe three of the adapted attacks that have been observed in MySpace: malware propagation, spam, and phishing.

Malware Propagation

Malware creators aim to spread their malicious content to as many victims as possible. In fact, over the past couple of years, MySpace was attacked by at least one instance of each of the following malware categories: worms, spyware, and adware. For the remainder of this section, we will detail the most interesting occurrences of these attacks, and we will explain the threats they pose to social networking communities.

Worms

The most successful example of rapid worm propagation in a social networking community occurred in MySpace on October 4, 2005. The worm was called the “Samy worm,” and it generated more than a million friend requests for its creator (Samy) over the course of a single day. The basic operation of this worm was quite simple but extremely clever (Ratkiewicz et al. 2011). First, Samy wrote the worm using Javascript, and then, he embedded it in his MySpace profile. MySpace disallows users from adding scripts to their profiles by filtering scripting tags and removing specific strings (e.g., “javascript”). To evade these filters, Samy exploited the behavior of popular web browsers such as Internet Explorer. Specifically, he hid the code inside a Cascading Style Sheet (CSS) tag and obfuscated the strings that MySpace would filter (e.g., “javascript” became “java\ascript”). Thus, even though MySpace had security mechanisms in place, the lax security of certain web browsers allowed the obfuscated code to execute successfully.

When a MySpace user accessed Samy’s profile, the embedded Javascript code sent Samy a friend request on that user’s behalf. The code also embedded itself in the user’s profile (in the same manner that it was embedded in Samy’s profile). Consequently, when other MySpace users visited the newly infected profile, the code would send Samy friend requests from those users and propagate itself to their profiles. As Samy described it, “If 5 people viewed my profile, that’s 5 new friends. If 5 people viewed each of their profiles, that’s 25 more new friends.”

Fortunately, this worm was relatively harmless for its infected users because it only affected their

MySpace profiles and not their actual machines. However, the same cannot be said for MySpace. Due to the worm’s viral growth pattern, the MySpace administrators were forced to temporarily shut down the site to stop the worm’s propagation and remove the worm’s code from the infected users’ profiles. The service outage was relatively brief, but this incident clearly illustrates the potential damage that social networking worms are capable of inflicting. Specifically, this worm teaches two very important lessons. First, the success of the worm’s obfuscated code clearly highlights the importance of web site security as well as web browser security. A delicate balance exists between functionality and security, and this balance must be respected when designing and developing online communities. Second, the worm’s propagation speed showcases how quickly the entire community could become infected. In this case, more than a million users were affected in less than 24 h.

A much more dangerous social networking worm appeared on MySpace in the middle of July 2006. Similar to the Samy worm, this worm propagated itself through the profiles of unsuspecting MySpace users. However, unlike the Samy worm, this worm’s code was hidden inside a malformed Shockwave Flash (.swf) file that directly exploited a vulnerability on users’ machines. Additionally, instead of generating innocuous friend requests, this worm actually redirected users’ web browsers to a politically charged blog posting.

The manner in which this worm spread is as follows. First, the worm’s creator generated a malformed .swf file and embedded it in a MySpace profile. This .swf file exploited a critical vulnerability in Macromedia Flash Player v8.0.24.0 and earlier versions, which allows an attacker to execute code on an affected machine. When a MySpace user with a vulnerable Macromedia Flash Player accessed the profile, the code in the .swf file automatically redirected the user’s browser to a blog posting that contained political propaganda. Finally, the code propagated itself by embedding a copy of the .swf file in the infected user’s profile.

This worm was also relatively harmless; however, it was far more troubling than the Samy worm because it actually exploited a vulnerability on users' computers, which allowed it to execute code. Fortunately, the worm's creator was only interested in spreading political ideals because nothing prevented the worm from installing any number of malicious utilities on its victims' machines. For example, the worm could have installed spyware or adware, and it could have even zombified the infected machine (i.e., compromise the computer and enlist it in a botnet (Boyd 2006)). Combine these frightening, yet completely realistic, scenarios with the viral propagation patterns of these worms, and it becomes immediately obvious how devastating worms could be in a social networking environment and why they must be prevented.

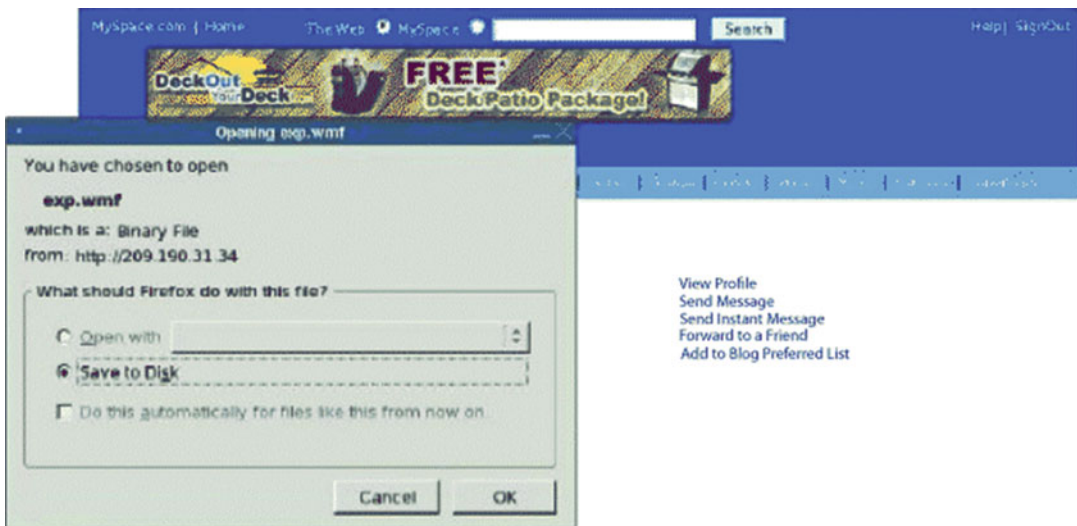
Spyware

Although spyware has yet to be distributed within the payload of a social networking worm, it has already made a few appearances in social networking communities. The most prominent example of spyware appearing in MySpace occurred towards the beginning of July 2006. At that time, an advertisement for deckoutyourdeck.com, which contained a malformed Windows Metafile

(.wmf) image, was inserted into one of the ad networks that MySpace uses. This malformed .wmf image exploited a critical vulnerability in the Graphics Rendering Engine of Windows, which allows remote code execution.

MySpace displays an ad banner, which is randomly selected from MySpace's supplying ad networks, at the top of every profile. When a user accessed a profile that displayed the deckoutyourdeck.com ad, the user was prompted to download the embedded .wmf image. Figure 2 shows a screenshot of this prompted download. If the user was running an unpatched version of Windows, the .wmf file installed an assortment of programs, including known spyware utilities such as PurityScan (Wang et al. 2011). These spyware utilities pose a serious threat to the user because they track the user's web browsing behaviors, and they install various other third-party applications without the user's consent.

Unfortunately, despite the fact that Microsoft released a patch for this vulnerability more than six months before the ad appeared on MySpace, over a million users were affected. Thus, this incident reveals two very important points. First, many users have a false sense of security in these communities. One of the most basic secure browsing principles is never to download questionable



Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities,
Fig. 2 Embedded .wmf image within a deckoutyourdeck.com advertisement

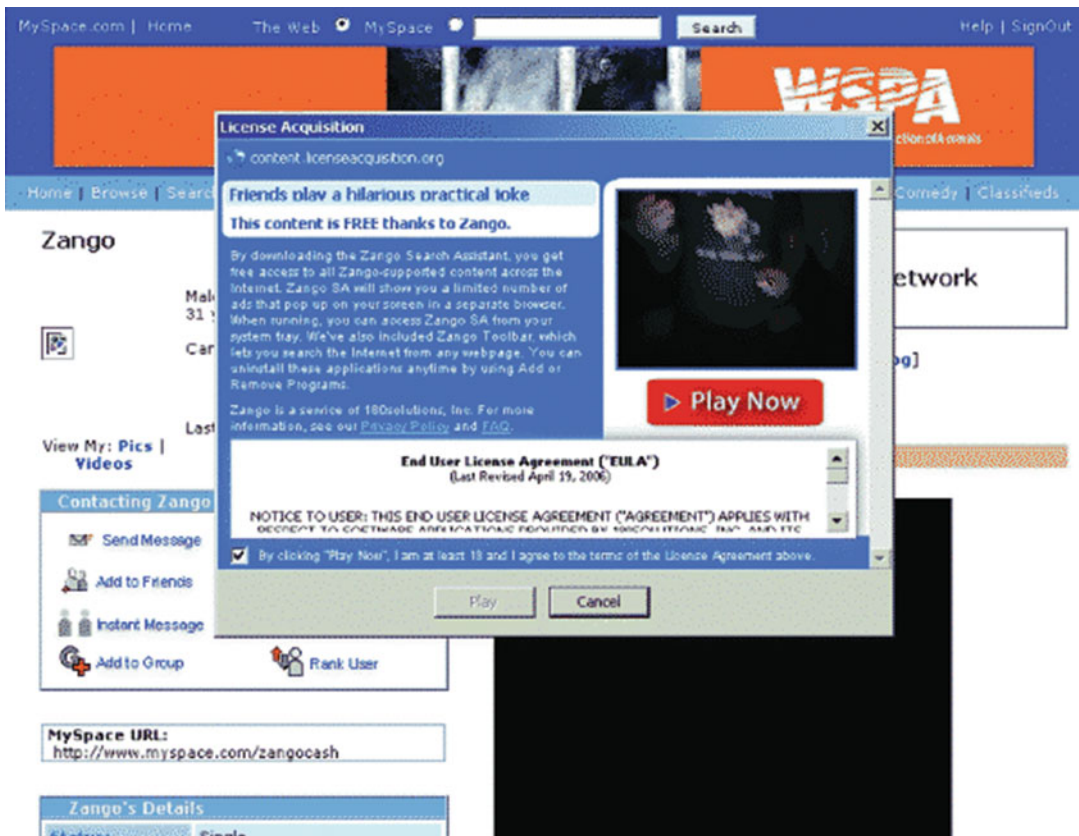
content, yet more than a million users gladly accepted this suspicious (and completely unsolicited) download request. Second, an alarming number of users are not vigilant about protecting themselves against security threats. This incident proves that at least a million users neglected to install security patches for more than 6 months. Therefore, social networking communities must assume that their users are completely vulnerable and take every precaution necessary to protect them from malicious content.

Adware

An interesting instance of adware appearing in MySpace occurred around the same time as the previous spyware example. A security researcher was browsing MySpace profiles and found two that were named after a known adware company called “Zango” (formerly known as “180solutions”).

Both profiles were created to deceive users into downloading and installing the Zango Search Assistant and Toolbar (two known adware programs) (Geer 2005). The first profile claimed that the programs could “protect kids from predators,” and the second profile was even more deceptive.

When a user accessed the second Zango profile, a popup with a license agreement immediately launched, asking the user to accept the license in order to play a video file. This license popup is shown in Fig. 3. If the user accepted the license by clicking the “Play Now” button, a video file began to play, but secretly, the Zango Search Assistant and Toolbar were also installed on the user’s system. As Fig. 3 illustrates, this popup contained a number of deceptive elements. First, the popup did not explicitly indicate that an installation was taking place. The text in the top-left corner of the popup mentions the Zango Search



Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities,
Fig. 3 Deceptive Zango popup license agreement

Assistant and Toolbar, but it does not tell the user that they will be installed. Additionally, none of the popup's buttons mention an installation. Instead, they are entitled "Play Now" and "Play," which implies that the only action will be the play-back of a video file. Finally, the most prominent features of the popup are the video's preview window in the top-right corner and the "Play Now" button. The actual license agreement and its preselected checkbox are positioned at the bottom, where they are likely to be overlooked. Consequently, many users played the video before they noticed the license agreement, and the adware was successfully installed on their machines.

According to Zango, both profiles were created by a Zango developer that was acting against the company's policy of not targeting MySpace. As a result, the deceptive video clip was removed, and the company released a public apology. However, this occurrence clearly illustrates the potential for abuse by adware companies in social networking communities. Additionally, the success of this deception further illustrates the naivety of users and the need for these communities to provide effective security mechanisms that protect against malicious content.

Spam

In addition to malware propagation, another type of attack on traditional communication media (e.g., email, blogs, instant messaging) is spam. Since MySpace provides similar communication facilities, they are also susceptible to spamming abuse. In fact, almost all spamming activities that occur outside the community can be recreated inside the community. To aggravate the problem, spammers can also use the profile information posted by users to target them outside the social networking community. Although MySpace's Terms of Use Agreement prohibits users from providing contact information such as email addresses and URLs on their profiles, many users still do so at their own peril. Consequently, spammers can use that information to spam those users using traditional techniques.

Spamming relies on the open nature of communications; thus, the vulnerabilities of

MySpace's communication facilities are proportionate to the openness of their access. On the conservative side, the commenting and bulletin systems are the most spam resistant because they can only be used by a user's friends. The blogging, IM, and friend request systems also provide the option of disallowing nonfriends from using them to contact a user, but this protection is not enabled by default. Thus, these three systems are more susceptible to spamming because any user can use them to contact other users that have not enabled this protection. The messaging system is the most vulnerable to abuse because it does not provide an effective spam prevention mechanism. It allows users to report a message as being spam, but those users have no assurances that immediate action will be taken. Therefore, users may receive a number of spam messages before MySpace administrators eliminate the offending spammer from the system. MySpace also provides a general block listing mechanism that allows users to completely prevent all communication from specific users. However, this feature is ineffective due to the ease with which users can create new MySpace profiles (i.e., if spammers have been blocked, they can create new profiles and continue their spamming activities undeterred).

Spammers abuse these communication systems for the same reason they abuse traditional communication facilities: promotion. Spammers want to expose their products, web sites, and viewpoints to as many individuals as possible, and spamming provides them with a technique to accomplish that goal. For social networking communities, this spamming activity represents a huge problem for a couple of reasons. First, spam content wastes a significant amount of resources, including storage space, bandwidth, and users' time. This last wasted resource leads us to the second major consequence of social network spam. When users waste time dealing with spam, it has a negative effect on their overall experience with these communities because it prevents them from participating in their desired activities. For example, when users are forced to sift through an inbox full of annoying spam messages, they are unable to send and receive messages. Similarly, when users are burdened with

removing an endless stream of spam comments on their blogs, they are unable to post new content. Consequently, if users become overwhelmed with spam, they will have no incentive to continue interacting with these communities, and the communities will be forced to shut down.

In addition to spamming the social networking community, spammers are also able to use the information on user profiles to more effectively spam users outside the community. By mining email addresses, IM screen names, and other contact information from these profiles, spammers can spam users using traditional techniques (e.g., email spam, spim, blog comment spam). Additionally, spammers can use the personal information on these profiles to construct user-specific spam content that is more relevant to the user, making it more likely to be read. For example, if a user's profile contains a great deal of sports-related information, a spammer can leverage this information to create a sports-oriented spam message for that user. Since the message's content matches the user's interests, the user is much more likely to read it, and as a result, the spam's sales pitch is more likely to be successful. To make their messages even more deceptive, spammers can also masquerade as users' friends. A user's profile contains a list of that user's friends; thus, spammers can use this information to make their messages appear as if they were sent by those friends. Since a user is much more likely to read and trust a message from a friend, the spammer has a higher probability of success with these disguised messages (Ross et al. 2005).

Phishing

In a traditional phishing attack, a phisher seeks to obtain a targeted user's sensitive information through means of deception. Historically, phishers have been interested in credit card information, banking information, and login information for various web sites (e.g., eBay, PayPal). As social networking communities have become more popular, complex networks of friends have been established within them. Consequently, social networking communities have become a prime target for phishers because they are portals to a large number of potential victims.

A MySpace phishing attack utilizes the same deceptive techniques that are employed in traditional phishing attacks. First, a user is presented with a seemingly legitimate URL (called a *phishing URL*) that appears to be affiliated with MySpace. This phishing URL is typically propagated in two distinct manners: the communication systems provided by MySpace and traditional communication channels (e.g., email, blogs, instant messaging). For example, in May 2006, a phisher used the MySpace spamming techniques described above to send a phishing message to various MySpace users. This message had "CHECK OUT these old school pictures . . ." as its subject and a phishing URL in its body (Technocrat 2006b). Upon accessing one of these phishing URLs, the user is directed to a fraudulent web page that appears identical to the authentic MySpace login page. When the user enters the necessary MySpace login information, the fraudulent page stores that information and uses it to redirect the user to the authentic MySpace community. Thus, the user is completely unaware of the attack, while the phisher successfully obtains the user's sensitive login information.

Phishing attacks represent a serious threat to MySpace users for three reasons. First, victimized users often lose control of their profiles. In many cases, phishers will immediately change the login information of profiles they have compromised, locking victims out of their own profiles. Since users spend a great deal of time and energy building new friendships, writing blogs, and customizing their profiles, it is somewhat traumatic when they lose control of their creations. Even in cases where a compromised profile's login information remains the same, the phisher's activities severely damage that profile's credibility. For example, if a phisher compromises a user's profile and begins spamming that user's friends, those friends will eventually distrust the compromised profile and remove it from their lists of friends. The second reason these attacks are dangerous is due to the viral propagation patterns that are possible in social networking communities. As shown with the malware examples above, one compromised user can quickly escalate into a network full of

compromised users. Specifically, once a single user's login information is compromised, the phisher can use that user's profile to propagate the attack to the user's friends. In our spam discussion, we mentioned a few MySpace communication systems that are somewhat spam resistant. However, the spam resistance of those systems assumes that a user's friends can be trusted (i.e., they are not spammers, phishers). Thus, if a user's profile becomes compromised, that user's friends are immediately at risk because the phisher can contact them under the guise of a profile they trust. As a result, a compromised user's friends are highly likely to become phishing victims (i.e., access the phishing URL), and by the time they realize they should distrust the compromised profile, it will be too late. The final threat posed by MySpace phishing attacks relies on the knowledge that many computer users reuse the same login information at many different sites (Lee et al. 2013). Thus, if a user's MySpace login information is compromised, that user's login information at those other sites is automatically compromised as well.

In addition to launching phishing attacks that specifically target social networking communities, phishers can also use the private information found in these communities to make their traditional phishing attacks more effective. As previously mentioned, many MySpace users include various pieces of personal information on their profiles (e.g., location, interests, occupation). Phishers can easily use this personal information to construct user-specific messages that a potential victim would be much more likely to read and trust. For example, a phisher could send the potential victim a fraudulent eBay email that contains auction information for products the victim would be interested in buying. Each of these products could be selected using the user's interests and associated with one or more phishing URLs. Since this email message contains user-specific content, the victim is more likely to read it and access one of its phishing URLs. As a result, this user-specific attack has a higher probability of success than a generic phishing attack.

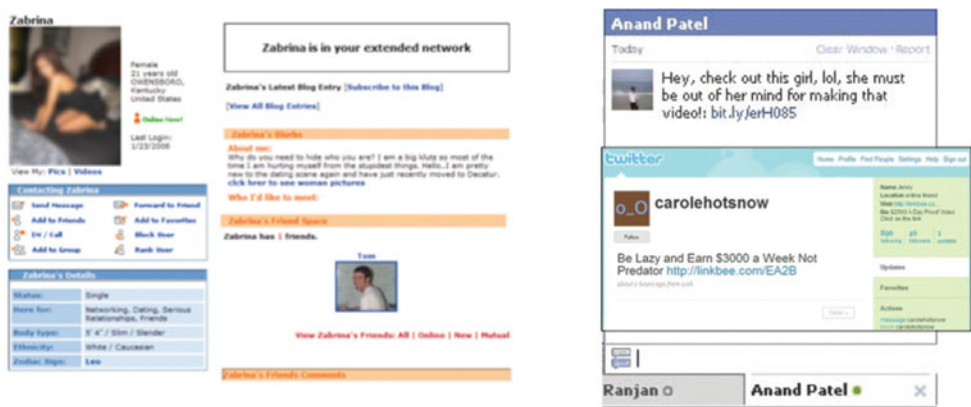
New Attacks Against Social Networking Communities

With networks like Facebook and Twitter attracting audiences of over 100 million users a month, they are becoming an important medium of communication. Unfortunately, the explosive growth of these communities and the conflicting interests of their various participants have generated a range of new attacks. In this section, we describe three types of emerging attacks: social spam, campaigns, and misinformation.

Social Spam

Social spam is "spam" on social networks. Such spam is a nuisance to people and hinders them from consuming information that is pertinent to them or that they are looking for (Irani et al. 2010; Wang et al. 2011). Besides being similar to email spam, it has various forms to present in social media such as user profile spam and web page spam. By employing text such as attractive words and multimedia such as pictures or videos, user profile spam aims to capture other users attention or clicks in the purpose of redirecting users to phishing or spam web pages or increasing account reputations by more clicks or followings. Moreover, web page spam mainly propagates false information such as phishing web page or low-quality information such as irrelevant advertisings to Internet users. Figure 4 shows the examples of social spam. Individual social networks are capable of filtering a significant amount of the spam they receive, although they usually require large amounts of resources (e.g., personnel) and incur a delay before detecting new types of spam.

Another type of social spam is called *collective attention spam*, which relies on users themselves to seek out the content where the spam will be encountered (Lee et al. 2013). Collective attention spammers have begun targeting information needs of these users and their curiosity about popular topics on social systems. They first get trending topics and keywords from various public information sources such as Google Trends and Twitter Trends in the real-time and postspam message with one of the topics or keywords so that the



(a) Example of user profile spam (b) Example of message spam



(c) Example of web page spam

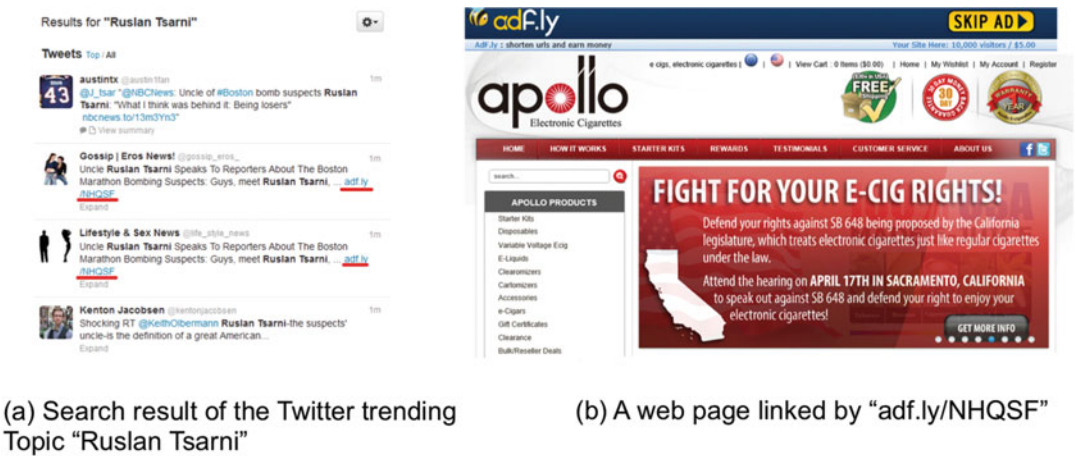
Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities, Fig. 4 Examples of social spam

message is exposed to the users who are interested in accessing the topic-related information. Figure 5a shows a search result of the Twitter trending topic “Ruslan Tsarni.” Two out of four most recent messages contain URLs linking to a web page in Fig. 5b which is an advertising page and irrelevant to the trending topic.

Campaigns

In the recent years, spammers have begun generating coordinated messages and creating coordinated campaigns. Researchers have observed various types of coordinated campaigns – political astroturfing campaigns (Ratkiewicz et al. 2011), fake reviews (Mukherjee et al. 2012), spam and

promotion campaigns (Lee et al. 2011). For examples, spammers for 2010 US Senate special election in Massachusetts (Mustafaraj and Metaxas 2010) created 9 Twitter accounts to generate 929 tweets during 138 min and sent these messages to users who previously mentioned about the election. All tweets have the identical signature: @account Message URL (e.g., @theRQ AG Coakley thinks Catholics shouldn't be in the ER, take action now! <http://bit.ly/8gDSP5>). Soon after, the recipients retweeted some of the messages 143 times which have reached to 61,732 Twitter users who are followers of the recipients. Note that in Twitter, if a user posts a message, his followers receive the message automatically. This



Granular Computing System Vulnerabilities: Exploring the Dark Side of Social Networking Communities,
Fig. 5 Spam messages targeting the Twitter trending topic “Ruslan Tsarni” and its linked web page

result presents that the attack using an astroturfing campaign in social media is powerful and successful.

As we have seen the above example, campaigns are more powerful than individual spam or attack because many spammers create coordinated contents during a short time period to target people. To make a group attack more effective, these spammers also take an advantage of crowdsourcing sites where they hire a number of workers whose job is to spread manipulated contents to the Web including social media. We often see these malicious tasks like “For \$5, I will submit two great reviews for your business” at Fiverr – one of popular crowdsourcing sites (Technocrat 2006b). A recent study reports that 90% of all tasks in crowdsourcing sites are for forming artificial campaigns, and workers have earned over multimillion dollars in total (Wang et al. 2012). These evidences indicate that campaigns are becoming a serious problem.

Misinformation

Popular online social networks such as Facebook and Twitter become one of major news sources in everyday lives for many people. Based on the trust relationships built among their users, OSNs obtain popularity from the convenience and efficiency of information dissemination and sharing in the end.

However, such trust relationships can possibly be exploited for spreading misinformation or rumors that may cause serious consequence such as the widespread panic in the general public (Nguyen et al. 2012). For example, In Twitter tweets, we observed the misinformation of swine flu at the outset of the large outbreak in 2009 (Coursey 2009) or the false report that President Obama was killed on the hacked Fox News’ Twitter feed widespreads in July 2011 (Greene 2011).

In addition, Twitter supports real-time propagation of information to a large group of users since tweets can be posted by a wide range of services: email, SMS text-messages, and Apps on smartphones. Therefore, Twitter provides an ideal environment for the dissemination of breaking-news directly from the news sources and/or geographical location of events (Castillo et al. 2011). For example, some users provide first-person observations or bring relevant knowledge from external sources into Twitter in an emergency situation (Vieweg 2010). This rapid process may not be able to separate true information from false rumors. Indeed, we observed that immediately after the 2010 earthquake in Chile in (Mendoza et al. 2010), when tweets contributed to increase the sense of chaos and insecurity in the local population.

Future Directions

In only a few years, social networking communities have made staggering strides in popularity. Aside from their social and economic impact, one of the most important questions is whether social networking environments are safe for their users. In this entry, we have analyzed this safety question and described several security and privacy threats that have translated into real attacks. Some of these attacks have been adapted from other Internet environments, including variants of malware propagation, spam, and phishing. Other attacks are new and unique to social networking environments, including social spam, campaigns, and misinformation. From our real world observations, it is clear that additional efforts should be made to protect social networking users. We hope that researchers will identify and counteract emerging variations of these known attacks and provide measures for safeguarding the future of these valuable communities.

Bibliography

- Boyd C (2006) Teenagers used to push Zango on Myspace. <http://www.vitalsecurity.org/2006/07/teenagers-used-to-push-zango-on.html>
- Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on World Wide Web, WWW '11, pp 675–684
- Coursey D (2009) Swine Flu Frenzy Demonstrates Twitter's Achilles Heel. http://www.pcworld.com/article/163920/swine_flu_twitter.html
- Geer D (2005) Malicious bots threaten network security. *IEEE Comput* 38(1):18–20
- Greene RA (2011) Secret service to probe Fox News. Twitter hack. http://www.cnn.com/2011/TECH/social.media/07/04/fox.hack/index.html?_s=PM:TECH
- Irani D, Webb S, Pu C (2010) Study of static classification of social spam profiles in myspace. In: Proceedings of the international AAAI conference on weblogs and social media
- Jagatic T et al (2007) Social phishing. *Commun ACM* 50(10):94–100
- Krebs B (2006) Hacked Ad seen on MySpace served spyware to a million. http://blog.washingtonpost.com/securityfix/2006/07/myspace_ad_served_adware_to_mo.html
- Lee K, Caverlee J, Cheng Z, Sui DZ (2011) Content-driven detection of campaigns in social media. In: Proceedings of ACM international conference on information and knowledge management
- Lee K, Kamath K, Caverlee J (2013) Combating threats to collective attention in social media: an evaluation. In: Proceedings of the international AAAI conference on weblogs and social media
- Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we rt? In: Proceedings of the first workshop on social media analytics, SOMA '10, pp 71–79
- Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the international World Wide Web conference
- Mustafaraj E, Metaxas P (2010) From obscurity to prominence in minutes: political speech and real-time search. In: Proceedings of the annual ACM web science conference
- Nguyen NP, Yan G, Thai MT, Eidenbenz S (2012) Containment of misinformation spread in online social networks. In: Proceedings of the annual ACM web science conference, WebSci '12, pp 213–222
- Ratkiewicz J, Conover M, Meiss M, Goncalves B, Flammini A, Menczer F (2011) Detecting and tracking political abuse in social media. In: Proceedings of the international AAAI conference on weblogs and social media
- Ross B et al (2005) Stronger password authentication using browser extensions. In: Proceedings of the 14th Usenix security symposium, pp 17–32
- Samy (2005) Technical explanation of the MySpace worm. <http://namb.la/popular/tech.html>
- Technocrat (2006a) Myspace phishing attacks on the rise. <http://djtechnocrat.blogspot.com/2006/05/myspace-phishing-attacks-on-rise.html>
- Technocrat (2006b) In a Race to Out-Rave, 5-Star Web Reviews Go for \$5. http://www.nytimes.com/2011/08/20/technology/finding-fake-reviews-online.html?_r=0
- Vieweg S (2010) Microblogged contributions to the emergency arena: discovery, interpretation and implications. In: Computer supported collaborative work
- Wang D, Irani D, Pu C (2011) A social-spam detection framework. In: Proceedings of the collaboration, electronic messaging, anti-abuse and spam conference
- Wang G, Wilson C, Zhao X, Zhu Y, Mohanlal M, Zheng H, Zhao BY (2012) Serf and turf: crowdurfing for fun and profit. In: Proceedings of the international World Wide Web conference

Granular Model for Data Mining

Anita Wasilewska¹ and Ernestina Menasalvas²

¹Computer Science Department, Stony Brook University, Stony Brook, NY, USA

²Departamento de Lenguajes y Sistemas Informaticos, Facultad de Informatica, Madrid, Spain

Article Outline

Glossary

Definition of the Subject

Introduction

Granular Model: Syntax and Semantics for Data Mining

Semantic Model

Granular Model Revisited: Satisfaction and Truth

Future Directions

Bibliography

Glossary

Data mining Data mining is a process that includes the following phases: creating the target data, data preprocessing, data mining proper, pattern evaluation, and knowledge presentation.

Syntax Syntax, or syntactical concepts, refer to simple relations among symbols and expressions of formal, symbolic languages. A symbolic language is a pair $\mathcal{L} = (\mathcal{A}, \mathcal{E})$, where $\mathcal{A}$ is an alphabet and $\mathcal{E}$ is the set of expressions of $\mathcal{L}$. The expressions are often the formulae of the language, but can also represent sets of sequences of formulae, sets of clauses, or other sets built out of symbols of the alphabet, or their subsets. The expressions of formal languages, even if created with a specific meaning in mind, do not carry themselves any meaning, they are just finite sequences of certain symbols. The meaning is

being assigned to them by establishing a proper semantics.

Semantics Semantics for a given symbolic language $\mathcal{L}$ assigns a specific interpretation in some domain to all symbols and expressions of the language. It also involves related ideas such as truth and model. They are called semantical concepts to distinguish them from the syntactical ones.

Model The word *model* is used in many situations and has many meanings but they all reflect some parts, if not all, of its following formal meaning. A structure M , called also an interpretation, is a model for a set $\mathcal{E}_0 \subseteq \mathcal{E}$ of expressions of $\mathcal{L}$ if and only if every expression $E \in \mathcal{E}_0$ is true in M .

Definition of the Subject

We present here three abstract models: descriptive, semantic, and granular. All of them are abstract structures that allow us to formalize some general properties of the data mining process and address the semantics-syntax duality inherent to any data mining process. A descriptive model formalizes the syntactical concepts and properties of the process. A semantic model formalizes its semantical properties. Finally, they form components of the granular model in which we establish a relationship between the descriptive and semantic models and provide a formal semantics for syntactical expressions of the language $\mathcal{L}$ of the descriptive model, thus justifying a use of the word model.

Moreover, we provide, within the granular model, a formal definition of data mining as the process of information generalization. The notion of generalization is defined in terms of granularity of steps of the process. In the model the data is represented in a form of knowledge systems. Each knowledge system has a granularity associated with it and the process changes, or not, its granularity. The notion of granularity is crucial for

defining some notions and components of the model, hence the granular model name.

Introduction

Data mining, as defined in 1996 by Piatetsky-Shapiro (Fayyad et al. 1996) is a step (crucial, but a step nevertheless) in a KDD (knowledge discovery in databases) process. The Piatetsky-Shapiro's definition states that the KDD process consists of the following steps: developing an understanding of the application domain; creating a target data set; choosing the data mining task i. e. deciding whether the goal of the KDD process is classification, regression, clustering, etc. . . .; choosing data preprocessing algorithms; choosing data mining algorithm(s); interpreting mined patterns; deciding if a re-iteration is needed; consolidating the discovered knowledge.

Since then the data mining (DM) term has evolved to become a name for all of the KDD process, or some parts of it, or even to be used as a name of an application of a data mining or learning algorithm.

In 1997 the Cross-Industry Standard Process for Data Mining (CRISP-DM) was proposed (Shearer 2000) to establish a standard for what they called, and others adopted, a data mining process. CRISP-DM standard was developed for business purposes and it included all of KDD process steps plus some extra steps such as a business understanding, business goal understanding followed by the KDD standard steps. Hence the KDD process became often a data mining process for industrial applications and was and is more and more often called just by the name of data mining.

To clarify these naming confusions we follow the standard terminology developed by data mining research in which we understand by data mining a KDD process in which its original data mining phase is now called *data mining proper phase*. For short we say that.

Data mining (DM) is a process that includes between the others the following phases: creating the target data, data preprocessing, data mining

proper, pattern evaluation, and knowledge presentation.

We formalize, within our granular model, the intuitive notions of the data mining preprocessing and data mining proper processes (section “[Data Mining Process](#)”).

One of the main goals of data mining is to *provide comprehensible descriptions* of information extracted from the data bases. The descriptions come in different forms. In the case of classification problems it might be a set of characteristic or discriminant rules, it might be a decision tree or a neural network with fixed set of weights. In the case of association analysis it is a set of associations (frequent itemsets), or association rules with accuracy parameters. In the case of cluster analysis it is a set of clusters, each of which has its own description and a cluster name. In the case of approximate classification by the rough set analysis it is usually a set of discriminant or characteristic rules (with or without accuracy parameters) or a set of decision tables.

If the number of descriptions is relatively small and all descriptions are of a reasonable length say that we obtained a generalized knowledge from the initial target database; that we mined more comprehensive, more general information. If an algorithm would reduce, for example, 100,000 records of the database of size 500 (number of attributes of the initial database) to 20 descriptions of size 10, we surely would say that we well generalized our information. Once it is done, a natural question of a *measure of quality* of such syntactic generalization and hence the validity of our method arises and this is being handled by our semantic model (section “[Semantic Model](#)”). The format of the descriptions is defined in our descriptive model (section “[Descriptive Model](#)”). The correctness and quality of syntactical descriptions is defined in our granular model by a satisfaction relation. The satisfaction relation (sections “[K- Satisfaction and K-Truth](#)” and “[Granular Model Revisited: Satisfaction and Truth](#)”) establishes a connection between syntactical and semantical approaches, and hence describe formally the syntax-semantic duality of the data mining process.

Granular Model: Syntax and Semantics for Data Mining

We usually view data mining results and present them to the user in their descriptive, i. e. syntactic form as it is the most natural form of communication. But the data mining process is deeply semantic in its nature. We hence build our granular model on two levels: syntactic and semantic. The syntactic level is represented by a descriptive model, and semantic level is described by the semantic model. These mantics-syntax duality of data mining process is expressed in our granular model by the satisfiability relation.

Moreover, we use a granular model to provide a formal definition of data mining as the process of information generalization (section “[Data Mining as Generalization](#)”). In the model the data preprocessing and data mining algorithms are defined as certain operators that act on data represented in a form of knowledge systems (Definition 9). Each knowledge system has a granularity associated with it (Definition 10) and the operators change, or not, its granularity. The notion of granularity is crucial for defining all notions and components of the model, hence the granular model name.

Granular Model Definition

For a given data mining application one defines in detail all the components of the model (see Examples 7, 14). We provide here, as a part of the granular model definition, a general form of such definitions.

Definition 1 A **granular model** is a system $\mathbf{GM} = (\mathcal{SM}, \mathcal{DM}, \models)$ where:

- $\mathcal{SM}$ is a **semantic model**;
- $\mathcal{DM}$ is a **descriptive model**;
- $\models \subseteq \mathcal{P}(U) \times \mathcal{E}$ is called a **satisfaction relation**, where U is the universe of $\mathcal{SM}$ and $\mathcal{E}$ is the set of descriptions defined by the $\mathcal{DM}$.

The satisfaction relation $\models$ (Definition 31) establishes the relationship between expressions of the semantic and descriptive models. It hence established formally the syntax-semantics duality

of the data mining process. The models are $\mathcal{SM}$ and $\mathcal{DM}$ are defined in Definitions 13 and 21, respectively.

Motivation and Examples

The semantic model is the most important component of our granular model. The intuitions behind the definition of the semantic model are as follows. When we perform the data mining procedures the first step in any of them is to drop the key attribute. This step allows us to introduce similarities in the database as records do not have anymore their unique identification. The input into the data mining process is hence always a data table obtained from the target data by removal of the key attribute. We call it a *target data table*.

As the next step we represent, following the rough set model, our target data table as Pawlak’s information system (Pawlak 1981) with the universe U by adding a new, non attribute column for the record names, i. e. objects of U . We take this set U as the universe of our model of $\mathcal{SM}$.

Data mining, as it is commonly said, is a process of generalization. In order to model this process we have first to define what does it mean from semantical point of view that one stage of the process is more general then the other. The idea behind is very simple. It is the same as saying that $(a + b)^2 = a^2 + 2ab + b^2$ is a more general formula then the formula $(2 + 3)^2 = 2^2 + 2 \cdot 2 \cdot 3 + 3^2$. This means that one description (formula) is more general then the other if it describes more objects. From a semantical point of view it means that data mining process consists of putting objects (records) in sets of objects. From a syntactical point of view the data mining process consists of building descriptions (in terms of attribute, values of attributes pairs) of these sets of objects, with some extra parameters, if needed.

To model a situation that allows us to talk about descriptions of sets of records we extend the notion of Pawlak’s model of an information system to our notion of a *knowledge system* (Definition 9). The universe the knowledge system consists of some subsets of U , i. e. elements of $\mathcal{P}(U)$. For example a target data table (after

preprocessing) and the corresponding representation by Pawlak's information system, and a knowledge system with universe U of *granularity one* are as follows.

Example 2 Target Data Table T_0 .

a_1	a_2	a_3
small	small	medium
medium	small	medium
small	small	medium
big	small	small
medium	medium	big
small	small	medium
big	small	small
medium	medium	big
small	small	medium
big	small	medium
medium	medium	small
small	small	medium
big	small	big
medium	medium	small

Target Information System I_0 .

U	a_1	a_2	a_3
x_1	small	small	medium
x_2	medium	small	medium
x_3	small	small	medium
x_4	big	small	small
x_5	medium	medium	big
x_6	small	small	medium
x_7	big	small	small
x_8	medium	medium	big
x_9	small	small	medium
x_{10}	big	small	medium
x_{11}	medium	medium	small
x_{12}	small	small	medium
x_{13}	big	small	big
x_{14}	medium	medium	small

A *knowledge system* corresponding to target table T_0 is a system that all its objects are one element sets $\{x\}$ for corresponding $x \in U$ of T_0 . We call such knowledge systems (Definition 11) a target knowledge system based on the target data table T_0 . The knowledge systems with the property that all its objects are one element sets are called systems of *granularity one*.

Example 3 Target Knowledge System K_0 .

$\mathcal{P}^1(U)$	a_1	a_2	a_3
$\{x_1\}$	small	small	medium
$\{x_2\}$	medium	small	medium
$\{x_3\}$	small	small	medium
$\{x_4\}$	big	small	small
$\{x_5\}$	medium	medium	big
$\{x_6\}$	small	small	medium
$\{x_7\}$	big	small	small
$\{x_8\}$	medium	medium	big
$\{x_9\}$	small	small	medium
$\{x_{10}\}$	big	small	medium
$\{x_{11}\}$	medium	medium	small
$\{x_{12}\}$	small	small	medium
$\{x_{13}\}$	big	small	big
$\{x_{14}\}$	medium	medium	small

Assume now that we have applied some algorithm ALG_1 and it has returned a following set $\mathcal{D} = \{D_1, D_2, \dots, D_7\}$ of descriptions (in the granular model, $D_i \in \mathcal{E}$ and $\mathcal{E}$ is defined in the descriptive model. We write s for the attribute value *small*, m for *medium* and b for *big*.

$$\begin{aligned}
 D_1 &: (a_1 = s) \cap (a_2 = s) \cap (a_3 = m), \\
 D_2 &: (a_1 = m) \cap (a_2 = s) \cap (a_3 = m), \\
 D_3 &: (a_1 = m) \cap (a_2 = m) \cap (a_3 = b), \\
 D_4 &: (a_1 = m) \cap (a_2 = m) \cap (a_3 = s), \\
 D_5 &: (a_1 = b) \cap (a_2 = s) \cap (a_3 = s), \\
 D_6 &: (a_1 = b) \cap (a_2 = s) \cap (a_3 = m), \\
 D_7 &: (a_1 = b) \cap (a_2 = s) \cap (a_3 = b).
 \end{aligned}$$

Now, a natural question arises: how well this set of descriptions describes our original data i. e. how accurate is the algorithm ALG_1 we have used to find them, how accurate is the *knowledge* we have thus obtained out of our data. To answer this, we use the target information system with the universe U , and for any $D \in \mathcal{D}$ we examine a set $S(D) = \{x \in U : D\}$ called a truth set for D . We define it formally in the Definition 27. Intuitively, the sets $S(D) = \{x \in U : D\}$ contain all records (i. e. their identifiers) with the same description given in terms of attribute, values of attribute pairs. The descriptions do not need to utilize all attributes of the target data, as it is often the case,

and one of ultimate goals of data mining is to find descriptions with as few attributes as possible.

In association analysis the descriptions can represent the frequent itemsets. For example, for a frequent three itemset $D = i_1 i_2 i_3$, the truth set $S(D)$ represents all transactions that contain items i_1, i_2, i_3 .

Descriptions come in different forms, depending on the data mining goal and application. We define formally a general form of descriptions as a part of the descriptive model (Definition 21).

For the data presented in the Examples 2, 3 and their descriptions $D_i \in \mathcal{D}$ the sets $S(D_i)$ are as follows.

$$\begin{aligned} S_1 &= S(D_1) = \{x \in U : D_1\} = \{x_1, x_3, x_6, x_9, x_{12}\}, \\ S_2 &= S(D_2) = \{x \in U : D_2\} = \{x_2\}, \\ S_3 &= S(D_3) = \{x \in U : D_3\} = \{x_5, x_8\}, \\ S_4 &= S(D_4) = \{x \in U : D_4\} = \{x_{11}, x_{14}\}, \\ S_5 &= S(D_5) = \{x \in U : D_5\} = \{x_4, x_7\}, \\ S_6 &= S(D_6) = \{x \in U : D_6\} = \{x_{10}\}, \\ S_7 &= S(D_7) = \{x \in U : D_7\} = \{x_{13}\}. \end{aligned}$$

We represent our results in a form of a *knowledge system* as follows. We write, as before, s for the attribute value *small*, m for *medium* and b for *big* and we put the names of proper subsets of U in the second representation of K_1 .

Example 4 Resulting Knowledge System K_1 .

$K_1(U)$	a_1	a_2	a_3
$\{x_1, x_3, x_6, x_9, x_{12}\}$	s	s	m
$\{x_2\}$	m	s	m
$\{x_5, x_8\}$	m	m	b
$\{x_{11}, x_{14}\}$	m	m	s
$\{x_4, x_7\}$	b	s	s
$\{x_{10}\}$	b	s	s
$\{x_{13}\}$	b	s	b

Resulting Knowledge System K_1 .

$K_1(U)$	a_1	a_2	$a_3 a$
S_1	s	s	m
S_2	m	s	m
S_3	m	m	b

(continued)

$K_1(U)$	a_1	a_2	$a_3 a$
S_4	m	m	s
S_5	b	s	s
S_6	b	s	s
S_7	b	s	b

The representation of data mining results in a form of a knowledge system allows us to define how good is the *knowledge* obtained by a given algorithm. In our case the knowledge obtained describes 100% of our target data as

$$S_1 \cup S_2 \cup S_3 \cap \dots \cup S_7 = \{x_1, x_2, \dots, x_{14}\} = U.$$

Observe that the sets $S_1, \dots, S_7$ are also disjoint and non-empty, i. e. they form a partition of the universe U . In such a case the knowledge system (and the algorithm) will be called *exact* (see Definition 12).

Moreover, we can see that the resulting system K_1 is *more general* than the input data, as represented by the target system K_0 because its *granularity* (i. e. maximum of cardinality of its granules, i. e. elements of its universe) is higher than the granularity of K_0 . This observation motivates the formal Definition 15. The granularity of our K_0 is one, as is the granularity of all target knowledge systems.

The *granularity* of K_1 is $\max\{|S_1|, \dots, |S_7|\} = \max\{5, 1, 2\} = 5$.

Let now assume that we have applied to our target data T (represented by K_0) another algorithm ALG_2 and it returned two descriptions D_1, D_2 under a condition that we need only descriptions of length 2, such that the corresponding attribute-value pairs appear in the target data with frequency $\geq 30\%$.

Example 5 Consider the following two descriptions D_1, D_2 .

- $D_1 : (a_1 = s) \cap (a_2 = s), D_2 : (a_2 = s) \cap (a_3 = m).$
- Now we evaluate: $S_1 = S(D_1) = \{x_1, x_3, x_6, x_9, x_{12}\},$
- $S_2 = S(D_2) = \{x_1, x_2, x_3, x_6, x_9, x_{10}, x_{12}\}.$

The description D_i fulfills the algorithm condition because it's of length 2 and $\frac{|S_1|}{|U|} = \frac{5}{14} \sim 0.3508$. The frequency% is greater then 30%.

The description D_2 fulfills the algorithm condition because it's of length 2 and $\frac{|S_1|}{|U|} = \frac{7}{14} = 0.5$. The frequency% is hence greater then 30%.

Incorporating the parameters of # of attributes in the description and their frequency imposed by the ALG_2 into our knowledge system we obtain the following table. The table is incomplete, as there are more descriptions fulfilling the algorithm conditions.

Example 6 Knowledge System K_2 .

$K_2(U)$	a_1	a_2	a_3	#of attr	frequency%
S_1	s	s	—	2	35
S_2	—	s	m	2	50

The sets S_1, S_2 do not form a partition of the universe U as $S_1 \cap S_2 \neq \emptyset$ and moreover, $S_1 \cup S_2 \neq U$. The algorithm ALG_2 is hence *not exact*. It describes only 57% of the target data and what is described is described following the frequency conditions. Of course K_2 is more general then K_0 and the algorithm ALG_2 generalized the target data, even if in an incomplete way. Now we form a new set of descriptions D_3, D_4 associated with the resulting knowledge system K_2 .

Example 7

- $D_3 : (a_1 = s) \cap (a_2 = s) \cap (\# \text{of attr} = 2) \cap (\text{frequency\%} = 36)$,
- $D_4 : (a_2 = s) \cap (a_3 = m) \cap (\# \text{of attr} = 2) \cap (\text{frequency\%} = 50)$.

We re-write D_3, D_4 as follows. $D_3 : D_1 \cap (\# \text{of attr} = 2) \cap (\text{frequency\%} = 36)$, $D_4 : D_2 \cap (\# \text{of attr} = 2) \cap (\text{frequency\%} = 50)$, where D_1, D_2 are descriptions considered in Example 5.

We say that the set $S_1 = S(D_1) = \{x \in U : D_1\}$ is a *truth set* for the description D_3 under restrictions defined by the descriptions $(\# \text{of attr} = 2) \cap (\text{frequency\%} = 36)$. The set $S_2 = S(D_2)$ is the truth set for D_4 under restrictions $(\# \text{of attr} = 2) \cap (\text{frequency\%} = 50)$.

Semantic Model

The definition of a semantic model presented here is a simpler and more comprehensible version of (Wasilewska et al. 1998; Wasilewska und Menasalvas 2004a; Wasilewska und Ruiz 2006, 2008). The initial investigations of the subject appeared also in (Hadjimichael und Wasilewska 1998; Martinez et al. 2002; Menasalvas et al. 2001, 2002). The notion of a knowledge system is central to the definition and development of the semantic model, as we have seen in section “[Motivation and Examples](#)”. We define it formally in the next section. We next use the knowledge system to formalize a notion of a granule and granularity (Definition 10) that is central to our granular model.

Knowledge Systems and Granularity

Observe that the table depicting the knowledge system K_2 (Example 6) is incomplete, i. e. not all attributes have assigned values. This fact in the formal definition will be represented by a *partial functioning*. Also K_2 contains some new *attributes describing properties of its granules*. In the formal definition we will call these new attributes the *knowledge attributes*.

The formal definitions of information system, knowledge and target knowledge systems, and their granularity and exactness are as follows.

Definition 8 A *Pawlak's information system* is a system $I = (U, A, V_A, f)$, where $U \neq \emptyset$ is called a *set of objects*; $A \neq \emptyset, V_A \neq \emptyset$ are called the *set of attributes and values of attributes*, respectively; $f : U \times A \rightarrow V_A$ is called an *information function*.

Definition 9 A *knowledge system* based on $I = (U, A, V_A, f)$ is a system

$$K = (K(U), A, E, V_A, V_E, g)$$

where:

- The **universe** $K(U)$ of K is a subset of the set $\mathcal{P}(U)$ of all subsets of the universe U of I , i. e. $K(U) \subseteq \mathcal{P}(U)$;

- E is a finite set of **knowledge attributes** (k -attributes) such that $A \cap E = \emptyset$;
- V_E is a finite set of **values of k - attributes**;
- g is a partial function called a **knowledge function** (k -function);
- $g : \mathcal{P} \times (A \cup E) \rightarrow (V_A \cup V_E)$ is such that:
 - (i) $g \mid (\cup_{x \in U} \{x\} \times A) = f$;
 - (ii) $\forall S \in \mathcal{P}, \forall a \in A ((S, a) \in \text{dom}(g) \Rightarrow g(S, a) \in V_A)$;
 - (iii) $\forall S \in \mathcal{P}, \forall e \in E ((S, e) \in \text{dom}(g) \Rightarrow g(S, e) \in V_E)$;

We use the above notion of a knowledge system to define the granules of the universe and the granularity of the system, and hence later, the granularity of the data mining process.

Definition 10 Any set $S \in \mathcal{P}(U)$ i. e. $S \subseteq U$ is called a **granule** of U . The cardinality $|S|$ of S is called a **granularity** of S . The set

$$Gr_K = \{S \in \mathcal{P} : \exists b \in (E \cup A) ((S, b) \in \text{dom}(g))\}$$

is called a **granule universe** of K . A number $gr_K = \max \{|S| : S \in Gr_K\}$ is called a **granularity** of K .

Observe that by conditions (ii), (iii) of Definition 9, $Gr_K = K(U)$ and it justifies its name. The condition (i) of Definition 9 says that when $E = \emptyset$, the k -function g is total on the set $\{\{x\} : x \in U\} \times A$ and $\forall x \in U \forall a \in A (g(\{x\}, a) = f(x, a))$. We denote $\mathcal{P}^1(U) = \{\{x\} : x \in U\}$.

Definition 11 Let $I = (U, A, V_A, f)$. Any system $K^1 = K^1 = (\mathcal{P}^1(U), A, \emptyset, V_A, \emptyset, g) = (\mathcal{P}^1(U), A, V_A, g) = (\mathcal{P}^1(U), A, V_A, g)$ is called a **target knowledge system** based on I .

Observe that any target knowledge system has granularity one. Finally, we define the *exactness* of a knowledge system as follows.

Definition 12 A knowledge system $K = (K(U), A, E, V_A, V_E, g)$ is called **exact** if all its granules Gr_K form a partition of the universe U .

The system K_2 from Example 4 is exact, the system K_3 from Example 6 is not exact.

In our model we view data mining algorithms as certain *operators*. For example, our ALG_1 is represented in the semantic model by an operator p_1 acting on some subset of a set $\mathcal{K}$ of knowledge systems, such that $p_1(K_0) = K_1$. ALG_2 is represented in the model by an operator p_2 also acting on some (may be different) subset of the set $\mathcal{K}$ of knowledge systems, such that $p_2(K_0) = K_2$. We put all the above observations into a formal notion of a semantic model.

Definition 13 A **semantic model** is a system $\mathbf{SM} = (\mathcal{P}(U), \mathcal{K}, \mathcal{G})$, where:

- $U \neq \emptyset$ is the **universe**;
- $\mathcal{K} \neq \emptyset$ is a set of knowledge systems, called also data mining process states;
- $\mathcal{G} \neq \emptyset$ is the set of **operators**;
- Each operator $p \in \mathcal{G}$ is a partial function on the set of all data mining process states, i. e. $p : \mathcal{K} \rightarrow \mathcal{K}$.

The semantic model is always being built for a given application. The target data is represented first in a form the target information system with the universe U , as in Example 2 and then in the form of target knowledge system as in the Example 3.

The semantic model based on our Examples 2, 3, and 4 are as follows.

Example 14 $\mathbf{SM}_1 = (\mathcal{P}(U), \mathcal{K}, \mathcal{G})$, where: $U = \{x_1, x_2, \dots, x_{14}\}$; $\mathcal{K} = \{K_0, K_1, K_2\}$; $\mathcal{G} = \{p_1, p_2\}$; Each $p_i \in \mathcal{G}$, ($i = 1, 2$) is a partial function $p_i : \mathcal{K}_1 \rightarrow \mathcal{K}_1$, such that $p_1(K_0) = K_1$, $p_2(K_0) = K_2$.

Data Mining as Generalization

In order to model within our semantic model data mining as a process of generalization we first introduce the following definition of *generalization relation* based on a notion of granularity.

Definition 15 A relation $\preceq \subseteq \mathcal{K} \times \mathcal{K}$ is called a **generalization relation** if the following condition holds for any $K, K' \in \mathcal{K}$.

$K \preccurlyeq K'$ if and only if $gr_K \leq gr_{K'}$,

where gr_K denotes the granularity of K (Definition 10).

Observe that for K_0, K_1, K_2 from Example 14, $gr_{K_0} = 1 \leq 5 = gr_{K_1} \leq 7 = gr_{K_2}$, and the system K_2 is the most general. But at the same time K_1 is exact and K_2 is not exact, so we have a trade off between exactness and generality.

As the next step we use the generalization relation to define the notion of a generalization operator as follows.

Definition 16 An operator $g \in \mathcal{G}$ is called a **generalization operator** if for any $K, K' \in \mathcal{K}$ such that $g(K) = K'$, we have that $K \preccurlyeq K'$.

Observe that both operators p_1, p_2 in Example 14 are generalization operators.

Data Mining Operators

In the data mining process the preprocessing and data mining are disjoint, inclusive/exclusive categories. The preprocessing is an integral and very important stage of the data mining process and needs as careful analysis as the data mining itself. Our framework allows us distinguish two disjoint classes of operators: the preprocessing operators $\mathcal{G}_{\text{prep}}$ and data mining proper operators $\mathcal{G}_{\text{dm}}$ and we put

$$\mathcal{G} = \mathcal{G}_{\text{prep}} \cup \mathcal{G}_{\text{dm}}.$$

The paper (Wasilewska et al. 2005) contains detailed formal definitions, their motivation, and discussion of these two classes. The model presented in Example 14 didn't include the preprocessing stage; it used the data mining proper operators only.

The main idea behind the concept of the operator is to capture not only the fact that data mining techniques generalize the data, but also to categorize existing methods. We want to do it in as exclusive/inclusive sense as possible. It means that we want to make sure that our categorization will be to distinguish as it should, for example

clustering from classification, making at the same time all classification methods fall into one category, called the classification operator while all clustering methods would fall into the category called the clustering operator. The third category is the association analysis described in our framework by the association operator. We don't include in our analysis purely statistical methods like regression, etc. . . . This gives us only three data mining classes of operators to consider: classification $\mathcal{G}_{\text{class}}$, clustering $\mathcal{G}_{\text{clust}}$, and association $\mathcal{G}_{\text{assoc}}$. The motivation, discussion and formal definition of these classes is included in the paper (Wasilewska und Ruiz 2008).

Theorem 17 Let $\mathcal{G}_{\text{class}}, \mathcal{G}_{\text{clust}}$ and $\mathcal{G}_{\text{assoc}}$ be the sets of all classification, clustering, and association operators, respectively. The following conditions hold.

- (1) $\mathcal{G}_{\text{class}} \neq \mathcal{G}_{\text{clust}} \neq \mathcal{G}_{\text{assoc}}$
- (2) $\mathcal{G}_{\text{assoc}} \cap \mathcal{G}_{\text{class}} = \emptyset$,
- (3) $\mathcal{G}_{\text{assoc}} \cap \mathcal{G}_{\text{clust}} = \emptyset$.

Data Mining Process

We model here two stages of the data mining process: preprocessing and data mining proper, as we were able to distinguish the data preprocessing and data mining proper operators as disjoint categories. We adopt the following definitions.

Definition 18 Any sequence $K_1, K_2, \dots, K_n$ ($n \geq 1$) of data mining states is called a **data preprocessing process**, if there is a preprocessing operator $G \in \mathcal{G}_{\text{prep}}$, such that $G(K_i) = K_{i+1}, i = 1, 2, \dots, n - 1$.

Definition 19 Any sequence $K_1, K_2, \dots, K_n$ ($n \geq 1$) of data mining states is called a **data mining proper process**, if there is a data mining proper operator $G \in \mathcal{G}_{\text{dm}}$, such that $G(K_i) = K_{i+1}, i = 1, 2, \dots, n - 1$.

The data mining process consists of the preprocessing process (that might be empty) and the data mining proper process. We know that the sets

$\mathcal{G}_{\text{prep}}$ and $\mathcal{G}_{\text{dm}}$ are disjoint. This justifies the following definition.

Definition 20 *A data mining process is any sequence $K_1, K_2, \dots, K_n$ ($n \geq 1$) of data mining states, such that $K_1, \dots, K_i$ ($0 \leq i \leq n$) is a preprocessing process and $K_{i+1}, \dots, K_n$ is a data mining proper process.*

Observe that in the semantic model $\mathcal{SM}_1$ (Example 14) we have only two processes: K_0, K_1 and K_0, K_2 . Both of them are data mining proper processes.

Descriptive Model

Given a semantic model $\mathcal{SM} = (\mathcal{P}(U), \mathcal{K}, \mathcal{G})$ We associate with it its descriptive counterpart defined below.

Definition 21 *A descriptive model is a system $\mathcal{DM} = (\mathcal{L}, \mathcal{E}, \mathcal{DK})$ where:*

- $\mathcal{L} = (\mathcal{A}, \mathcal{E})$ is called a **descriptive language**;
- $\mathcal{A}$ is a countably infinite set called the **alphabet**;
- $\mathcal{E} \neq \emptyset$ and $\mathcal{E} \subseteq \mathcal{A}^*$ is the set of **descriptive expressions** of $\mathcal{L}$;
- $\mathcal{DK} \neq \emptyset$ and $\mathcal{DK} \subseteq \mathcal{P}(\mathcal{E})$ is a set of **descriptions of knowledge states**.

As in the case of a semantic model, we build the descriptive model for a given application. We define here only a general form of the model. We assume however, that whatever the application is, the descriptions are always built in terms of attributes and values of the attributes, some logical connectives, some predicates and some extra parameters, if needed. For example, a neural network with its nodes and weights can be seen as a formal description (in an appropriate descriptive language), and the knowledge states would represent changes in parameters during the neural network training process, or a final, converged network.

The model we build here is a model for what we call a **descriptive data mining**, i. e. the data

mining for which the goal of the data mining process is to produce a set of descriptions in a language easily comprehensible to the user. For that reason we identify, for example, a decision tree constructed by the classification by the decision tree algorithm with the set of discriminant rules obtained from the tree.

In our Examples 2–14 we have used descriptions in the form $(a = v)$ to denote that the attribute a has a value v , but one might also use, like it is often done a predicate form $a(v)$ or $a(x, v)$ instead.

We define the components of $\mathcal{DM}$ in the following stages.

- | | |
|-------------|---|
| Stage 1 | For each $K \in \mathcal{K}$, we define (section “K-Descriptive Language”) its own descriptive language $\mathcal{L}_K = (\mathcal{A}_K, \mathcal{E}_K)$. |
| Stage 2 | For each $K \in \mathcal{K}$, and descriptive expression $F \in \mathcal{E}_K$, we define what does it mean that D satisfied in K ; i. e. we define (Definition 25) a satisfaction relation $\models_K$. |
| Stage 3 | For each $K \in \mathcal{K}$, and descriptive expression $F \in \mathcal{E}_K$, we define what does it mean that D is true K , i. e. $\models_K D$ (Definition 28). |
| Stage 4 | For each $K \in \mathcal{K}$, we define its set $\mathcal{DK} \subseteq \mathcal{P}(\mathcal{E}_K)$ of descriptions of its own knowledge (Definition 26). |
| Stage 5 | We use the languages $\mathcal{L}_K$ to define the descriptive language $\mathcal{L}$ (Definition 29). |
| Stage 6 | We use the descriptive expressions $\mathcal{E}_K$ of $\mathcal{L}_K$ to define the set $\mathcal{E}$ of descriptive expressions of $\mathcal{L}$ (Definition 30). |
| Later Stage | We use the satisfaction relations $\models_K$ to define the satisfaction relation $\models$ of our granular model $\mathbf{GM}$ (Definitions 1 and 31, respectively). |

K-Descriptive Language

As Stage 1 of our construction of the descriptive model, we define for each $K \in \mathcal{K}$, its own description language $\mathcal{L}_K$. The language depends on the semantic model and the goal of the data mining process. As we have said, the descriptions produced by data mining algorithms come in different forms. We build here a model for a **descriptive**

data mining, i. e. we assume that the descriptions are built from attributes and values of attributes and two logical connectives: conjunction (see descriptions associated with Examples 2–14) and implication. The implication connective is needed to model the different kind of rules that are being mined by data mining algorithms: discriminant and characteristic rules in classification analysis; association rules by association analysis; or other rules obtained by hybrid systems.

Definition 22 For any $K \in \mathcal{K}$, of $SM, K = (\mathcal{P}(U), A, E, V_A, V_E, g)$, we define the **descriptive language** of K , $\mathcal{L}_K$ as

$$\mathcal{L}_K = (\mathcal{A}_K, \mathcal{E}_K),$$

where $\mathcal{A}_K$ is called an **alphabet**, $\mathcal{E}_K$ the set of **descriptive expressions** such that

$$\mathcal{E}_K = \mathcal{D}_K \cup \mathcal{F}_K,$$

for $\mathcal{D}_K$ the set of **descriptive formulae** and $\mathcal{F}_K$ the set of **formulae** of $\mathcal{L}_K$ (Definition 23).

The alphabet $\mathcal{A}_K = VAR_K \cup \{\cap, \Rightarrow\} \cup \{(\cdot, \cdot)\}$, where $\{\cap, \Rightarrow\}$ is the set of logical connectives of $\mathcal{L}_K$. The set VAR_K of variables of K is also called its set of **atomic descriptions**. We put $VAR_K = \mathcal{D}_A \cup \mathcal{D}_E$, where $\mathcal{D}_A = \{(a = v) : a \in A, v \in V_A\}$, $\mathcal{D}_E = \{(a = v) : a \in E, v \in V_E\}$.

Atomic descriptions, i. e. the elements of $VAR_K = \mathcal{D}_A \cup \mathcal{D}_E$ represent minimal blocks of semantical description. Elements of $\mathcal{D}_A$ are atomic descriptions of minimal blocks built with use of the attributes of the initial target database. Elements of $\mathcal{D}_E$ are atomic descriptions of minimal blocks built with use of knowledge attributes used (if any) during the process of data mining. We use them to define the sets of descriptions and formulae as follows.

Definition 23 The set $\mathcal{D}_K$ of all **descriptive formulae** of $\mathcal{L}_K$ is the set

$$\mathcal{D}_K = \mathcal{A}\mathcal{D}_K \cup \mathcal{E}\mathcal{D}_K \cup \mathcal{F}_K,$$

where $\mathcal{A}\mathcal{D}_K, \mathcal{E}\mathcal{D}_K, \mathcal{F}_K$ are defined as below.

The set $\mathcal{A}\mathcal{D}_K \subseteq \mathcal{A}_K^*$ is called the set of **data attribute descriptions** and is the smallest set such that the following conditions hold.

- (1) $\mathcal{D}_A \subseteq \mathcal{A}\mathcal{D}_K$ (data attribute **atomic** description);
- (2) If $D_1, D_2 \in \mathcal{A}\mathcal{D}_K$, then $D_1 \cap D_2 \in \mathcal{A}\mathcal{D}_K$.

The set $\mathcal{E}\mathcal{D}_K \subseteq \mathcal{A}_K^*$ is called the set of **knowledge attribute descriptions** and is the smallest set such that the following conditions hold.

- (1) $\mathcal{D}_E \subseteq \mathcal{E}\mathcal{D}_K$ (knowledge attribute **atomic** description);
- (2) If $D_1, D_2 \in \mathcal{E}\mathcal{D}_K$, then $D_1 \cap D_2 \in \mathcal{E}\mathcal{D}_K$.

We distinguish two categories of formulae: one $\mathcal{F}_A$ describes a certain knowledge, the other, $\mathcal{F}_E$, the uncertain, or approximate, knowledge. The certain knowledge is expressed in our model in terms of attributes and values of attributes of the initial, target data only. The description of the approximate knowledge includes the extra parameters, describing properties of granules as defined by the knowledge attributes of K (Definition 9).

Definition 24 The set $\mathcal{F}_K$ of **formulae** of $\mathcal{L}_K$ is a union of two sets of formulae; $\mathcal{F}_A, \mathcal{F}_E$, i. e.

$$\mathcal{F}_K = \mathcal{F}_A \cup \mathcal{F}_E,$$

where $\mathcal{F}_A = \mathcal{A}\mathcal{D}_K \cup \mathcal{A}\mathcal{R}_K$ and $\mathcal{F}_E = \mathcal{A}\mathcal{E}\mathcal{D}_K \cup \mathcal{E}\mathcal{R}_K$ are defined below.

The set $\mathcal{A}\mathcal{E}\mathcal{D}_K = \{D_1 \cap D_2 : D_1 \in \mathcal{A}\mathcal{D}_K, D_2 \in \mathcal{E}\mathcal{D}_K\}$ is the set of **knowledge description formulae**, with knowledge attribute descriptions depicting uncertainty measures. The set $\mathcal{A}\mathcal{R}_K = \{(D_1 \Rightarrow D_2) : D_1, D_2 \in \mathcal{D}_A\}$ is the set of **attribute rule-formulae** depicting certain rules obtained during the data mining process. The set $\mathcal{E}\mathcal{R}_K = \{(D_1 \Rightarrow D_2) \cap D_3 : D_1, D_2 \in \mathcal{D}_A, D_3 \in \mathcal{D}_E\}$ is the set of **knowledge rule-formulae** depicting rules with uncertainty measures described obtained by descriptions from $\mathcal{D}_E$.

The formulae of the language $\mathcal{L}_K$ are **not** yet the data mining **rules**. They only describe a

syntactical form of the rules appropriate for a given application. Formulae from $\mathcal{F}_K$ become **rules determined by K** only when they do relate semantically to K , i. e. reflect the properties of our initial target database. In this case we say that they are **true**, or **true under the measures** described by descriptions from $\mathcal{D}_E$. We define, in the next section the notion of truthfulness, as we always do, via notion of satisfiability.

K- Satisfaction and K- Truth

In this section we define (Definition 25) a satisfaction relation $\models_K \subseteq \mathcal{P}(U) \times \mathcal{E}_K$ that establishes the relationship between descriptive expressions of the $\mathcal{L}_K$ and what they semantically represent in K . For any $(S, F) \in \models_K$ we say that S **satisfies F in K** and write it symbolically as

$$S \models_K F.$$

We call our satisfaction relation a k -satisfaction, as it represents a satisfaction relation relative to the system K . As the next step we define the notion of K -truth (Definition 28) i. e. we define what does it mean that a descriptive expression F is **true in K** , symbolically expressed as

$$\models_K F.$$

Let $K = (K(U), A, E, V_A, V_E, g)$ and let $\mathcal{L}_K = (\mathcal{A}_K, \mathcal{E}_K)$ be the description language defined by K (Definition 22), where $\mathcal{E}_K = \mathcal{D}_K \cup \mathcal{F}_K$.

Definition 25 A k -satisfaction relation $\models_K \subseteq \mathcal{P}(U) \times \mathcal{E}_K$ is defined by induction over the level of complexity of any descriptive expression $F \in \mathcal{E}_K$ as follows.

1. $F \in \mathcal{D}_K$.
 - (i) Let $(a = v) \in \mathcal{D}_A$ be an **atomic** attribute description. We define, for any $S \in \mathcal{P}(U)$, for any $a \in A, v \in V_A, S \models_K(a = v)$ if and only if $(S, a) \in \text{dom}(g)$ and $g(S, a) = v$;
 - (ii) Let $(e = v) \in \mathcal{D}_E$ is an **atomic** knowledge attribute description. We define, for any $S \in \mathcal{P}(U)$, for any $e \in E, v \in V_E, S \models_K(a = v)$ if and only if $(S, e) \in \text{dom}(g)$ and $g(S, e) = v$;

2. $F \in \mathcal{F}_K$.

We extend $\models_K$ to the set $\mathcal{F}_A = \mathcal{A}\mathcal{D}_K \cup \mathcal{A}\mathcal{R}_K$

- (iii) Let $D \in \mathcal{A}\mathcal{D}_K$ and $D = D_1 \cap \dots \cap D_n$. We define, for any $S \in \mathcal{P}(U)$, $S \models_K D$ if and only if $\forall (1 \leq i \leq n)(S \models_K D_i)$;
- (iv) Let $(D_1 \Rightarrow D_2) \in \mathcal{A}\mathcal{R}_K$, i. e. $D_1, D_2 \in \mathcal{A}\mathcal{D}_K$. We define, for any $S \in \mathcal{P}(U)$, $S \models_K(D_1 \Rightarrow D_2)$ if and only if $S \models_K D_1$ and $S \models_K D_2$;
- (v) Let $(D_1 \Rightarrow D_2) \cap D_3 \in \mathcal{E}\mathcal{R}_K$, i. e. $(D_1 \Rightarrow D_2) \in \mathcal{A}\mathcal{R}_K$ and $D_3 \in \mathcal{D}_E$. We define, for any $S \in \mathcal{P}(U)$, $S \models_K(D_1 \Rightarrow D_2) \cap D_3$ if and only if $S \models_K(D_1 \Rightarrow D_2)$ and $S \models_K D_3$.

This ends Stage 2 of the definition of the descriptive model and we are ready for Stage 3, i. e. to define the set of all descriptions of knowledge states.

Definition 26 Let $K \in \mathcal{K}$, the set $\mathcal{D}K \subseteq \mathcal{P}(\mathcal{E}_K)$ of **descriptions of knowledge states** of K is defined as follows

$$\mathcal{D}K = \{F \in \mathcal{E}_K : \exists S \in \text{Gr}_K(S \models_K F)\}.$$

Now we are ready to define a notion of F **true in K** , for any formula $F \in \mathcal{F}_K$, symbolically expressed by $\models_K F$. This notion relates the satisfaction relation $\models_K$, i. e. satisfiability of formulae in the system K with the initial target data as out ultimate point of reference. This connection is being established by the notion of the *truth set* defined below.

Definition 27 For any attribute data description $D \in \mathcal{A}\mathcal{D}_K \cup \mathcal{E}\mathcal{D}_K$ the set $S(D) = \{x \in U : D\}$ is called a **truth set** for D .

We define $\models_K F$ by induction over the level of complexity of any descriptive expression $F \in \mathcal{F}_K$ as follows.

Definition 28

1. $F \in \mathcal{A}\mathcal{D}_K$.

- (i) For any attribute data description $D \in \mathcal{AD}_K$, we define: $\models_K D$ if and only if $S(D) \models_K D$.
- (ii) For any attribute description formula $D \in \mathcal{AD}_K$, we put $\models_K D$ if and only if $S(D) \models_K D$.
- 2. $F \in \mathcal{F}_A$.
 - (iii) For an attribute rule-formula $(D_1 \Rightarrow D_2) \in \mathcal{AR}_K$, we define $\models_K(D_1 \Rightarrow D_2)$ if and only if $S(D_1) \models_K D_1$, $S(D_1) \models_K D_2$, and $S(D_1) \subseteq S(D_2)$ or $(D_1) \cap S(D_2) \neq \emptyset$.
- 3. $F \in \mathcal{AED}_K$.
 - (iv) For any knowledge attribute description $D_1 \cap D_2 \in \mathcal{AED}_K$, $D_1 \in \mathcal{AD}_K$, $D_2 \in \mathcal{ED}_K$ we define: $\models_K D_1 \cap D_2$ if and only if $S(D_1) \cap S(D_2) \models_K D_1$.
- 4. $F \in \mathcal{F}_E$.
 - (v) For an attribute rule-formula $(D_1 \Rightarrow D_2) \cap D_3 \in \mathcal{AER}_K$, we define $\models_K(D_1 \Rightarrow D_2) \cap D_3$ if and only if $S(D_1) \cap S(D_3) \models_K D_1$, $S(D_1) \cap S(D_3) \models_K D_2$, and $S(D_1) \cap S(D_3) \subseteq S(D_2) \cap S(D_3)$ or $(D_1) \cap S(D_2) \cap S(D_3) \neq \emptyset$.

In the case when $F \in \mathcal{F}_E$ and $\models_K F$ we say that F is **true in K under the measures D** , for a certain $D \in \mathcal{AED}_K$.

If $\models_K(D_1 \Rightarrow D_2)$ and the condition $S(D_1) \subseteq S(D_2)$ holds, then we call the formula $(D_1 \Rightarrow D_2)$ **C-true in K** , or **C-true under the measures D** , if $\models_K(D_1 \Rightarrow D_2) \cap D$.

If $\models_K(D_1 \Rightarrow D_2)$ and the condition $(D_1) \cap S(D_2) \neq \emptyset$ holds, then we call the formula $(D_1 \Rightarrow D_2)$ **D-true in K** , or **D-true under the measures D** , if $\models_K(D_1 \Rightarrow D_2) \cap D$.

The notion of D-truth reflects the semantics needed to define the discriminant rules for the classification analysis, and C-truth is needed for the characteristic rules, hence the names. As we have said before, the knowledge attributes from the set E describe uncertainty measures for the granules $S \in K(U)$. The formulae that incorporate the knowledge attribute descriptions $D \in \mathcal{ED}_K$ can be only *true* in K under some uncertainty measures (Example 5).

Descriptive Language and Descriptive Model

We have already constructed (section “[Descriptive Language and Descriptive Model](#)”) all sub-components of the Definition 21 of the descriptive model $\mathcal{DM}$ and now we proceed to complete it as follows.

Definition 29 We define the language of $\mathcal{DM}$ as $\mathcal{L} = (\mathcal{A}, \mathcal{E})$, where $\mathcal{A} = \cup\{\mathcal{A}_K : K \in \mathcal{K}\}$, and $\mathcal{E} = \cup\{\mathcal{E}_K : K \in \mathcal{K}\}$.

The set $\mathcal{DK}$ of all descriptions of knowledge states of the semantic model $\mathcal{SM} = (\mathcal{P}(U), \mathcal{K}, \mathcal{G})$ is the following.

Definition 30 Let $\mathcal{DK}$ be the set of descriptions of knowledge states of K (Definition 26). The set $\mathcal{DK} = \cup\{\mathcal{D}_K : K \in \mathcal{K}\}$ is a set of **descriptions of knowledge states** of $\mathcal{SM}$.

This completes the definition of the descriptive model as a system $\mathcal{DM} = (\mathcal{L}, \mathcal{E}, \mathcal{DK})$ where: $\mathcal{L} = (\mathcal{A}, \mathcal{E})$ is a descriptive language; $\mathcal{A}$ an alphabet and $\mathcal{E}$ is the set of descriptive expressions (Definition 29); $\mathcal{DK}$ is a set of descriptions of knowledge states (Definition 30).

Granular Model Revisited: Satisfaction and Truth

The granular model is defined (section “[Granular Model: Syntax and Semantics for Data Mining](#)”) as a system $\mathbf{GM} = (\mathcal{SM}, \mathcal{DM}, \models)$ where: $\mathcal{SM}$ is a semantic model as defined in the Definition 13; $\mathcal{DM}$ is a descriptive model (Definition 21); and $\models \subseteq \mathcal{P}(U) \times \mathcal{E}$ is a satisfaction relation. All the components of $\mathbf{GM}$ except the satisfaction relation have already been defined. As the last step we define the satisfaction relation and the notion of truth in $\mathbf{GM}$ as follows.

Definition 31 For any $S \in \mathcal{P}(U)$ and for any $F \in \mathcal{E}$, $S \models F$ if and only if $\exists K \in \mathcal{K}(S \models_K F)$.

Definition 32 We say that $F \in \mathcal{E}$ is **true in $\mathbf{GM}$** (symbolically $\models F$) if and only if $\exists K \in \mathcal{K}(\models_K F)$.

Future Directions

The models presented here set a general framework for future foundational investigations. They can be carried on three levels. One, the most general, would deal with further developments within the granular model. The second, a more specific one, would deal with applications of the methods and language developed within the granular model to specific domains. On this level one would build semantic, descriptive, and granular models for different data mining domains; i. e., build and examine specific models for classification, clustering, and for association analysis. Finally, the most specific third level would deal with building models for basic descriptive data mining algorithms: decision trees and rough sets classification, association and classification by association, and clustering. The general methodology of analysis of the data mining process developed for the granular model can hence serve as unifying language in which different data mining domains and algorithms can be studied, discussed, and compared.

Bibliography

- Chapman P (2000) CRISP-DM 1.0 – Step-by-step Data Mining CRISP-DM Consortium: 1.0
- Greco S, Matarazzo B, Slowinski R, Stefanowski J (2002) Importance and interaction of conditions in decision rules. In: Proceedings of Third International Conference RSCTC'02, Malvern, USA. Lecture Notes in Artificial Intelligence, vol 2475. Springer, Berlin, pp 255–262
- Hadjimichael M, Wasilewska A (1998) A hierarchical model for information generalization. In: Proceedings of the 4th joint conference on Information Sciences, Rough Sets, Data Mining and Granular Computing (RSDMG'98) North Carolina, vol II, pp. 306–309
- Han J, Kamber M (2000) Data mining: concepts and techniques. Morgan Kaufman
- Inuiguchi M, Tanino T (2003) Classification versus approximation oriented generalization of rough sets. Bull Int Rough Set Soc 7(1/2):19–25
- Komorowski J (2002) Modelling biological phenomena with rough sets. In: Proceedings of third international conference RSCTC'02, Malvern, USA. Lecture Notes in Artificial Intelligence, vol 2575. Springer, Berlin, p 13
- Lin TY (2002) Database mining on derived attributes. In: Proceedings of third international conference RSCTC'02, Malvern, USA. Lecture Notes in Artificial Intelligence, vol 2575. Springer, Berlin, pp 14–32
- Martinez FJ, Menasalvas E, Wasilewska A, Fernández C, Hadjimichael M (2002) Extension of relational management system with data mining capabilities. In: Proceedings of third international conference RSCTC'02, Malvern, USA. Lecture Notes in Artificial Intelligence, vol 2475. Springer, Berlin, pp 421–424
- Menasalvas E, Wasilewska E, Fernández C (2001) The lattice structure of the KDD process: mathematical expression of the model and its operators. Fundamenta Informaticae: Special Issue Int J Info Syst 48–62
- Menasalvas E, Wasilewska A, Fernández C, Martínez FJ (2002) Datamining – a semantical model. In: Proceedings of 2002 world congress on Computational Intelligence, Honolulu, pp. 435–441
- Pawlak Z (1981) Information systems. The or Found Inf Syst 6:205–218
- Pawlak Z (1991) Rough sets – theoretical aspects reasoning about data. Kluwer, Norwell
- Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery: an overview. In: Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in knowledge discovery and data mining. AAAI Press, Menlo Park, pp 1–34
- Polkowski L (2002) Rough sets – mathematical foundations. Physica, Heidelberg
- Shearer C (2000) The CRISP-DM model: the new blueprint for data mining. J Data Warehous 5(4):13–22
- Skowron A (1993) Data filtration: a rough set approach. In: Proceedings of rough sets, fuzzy sets and knowledge discovery. RSKD, pp. 108–118
- Wasilewska A, Ruiz EM, Fernández-Baizan MC (1998) A model for RSMD implementation. 1st international conference on Rough Sets and Current Trends in Computing (RSCTC'98), 22–26 June, Warsaw, pp 186–193
- Wasilewska A, Menasalvas E (2004a) Data preprocessing and data mining as generalization process. In: Proceedings of ICDM'04, the fourth IEEE International Conference on Data Mining, Brighton, 1–4 Nov, pp 133–137
- Wasilewska A, Menasalvas E (2004b) Data mining operators. In: Proceedings of ICDM'04, the fourth IEEE International Conference on Data Mining, Brighton, pp. 209–214
- Wasilewska A, Menasalvas E, Scharff C (2005) Uniform model for data mining. In: Proceedings of FDM05 (Foundations of Data Mining). In: ICDM2005, Fifth IEEE International Conference on Data Mining. Austin, Texas, pp. 19–27
- Wasilewska A, Ruiz EM (2006) Data mining as generalization: a formal model. In: Lin TY, Ohsuga S, Liau CJ, Hu X (eds) Foundations and novel approaches in data mining, Studies in computational intelligence, vol 9. Springer, Berlin, pp 99–126
- Wasilewska A, Ruiz EM (2008) Data preprocessing and data mining as generalization. In: Lin TY, Xie Y, Wasilewska A, Liau C-J (eds) Data mining:

- foundations and practice. Springer, Studies in computational intelligence, pp 453–468
- Wasilewska A, Ruiz EM (2005) A classification model: Syntax and semantics for classification. In: Proceedings of 10th international RSFDGrC 2005 conference, Regina, Canada, August/September 2005. Lecture notes in artificial intelligence, vol 2. Springer, Berlin, pp. 59–68
- Ziarko W, Fei X (2002) VPRSM approach to WEB searching. In: Proceedings of third international RSCTC'02 Conference, Malvern, USA, October 2002. Lecture notes in artificial intelligence, vol 2475. Springer, Berlin, pp. 514–522
- Ziarko W (1993) Variable precision rough set model. *J Comput Syst Sci* 46(1):39–59
- Yao JT, Yao YY (2002) Induction of classification rules by granular computing. In: Proceedings of third international RSCTC'02 Conference, Malvern, USA, October 2002. Lecture notes in artificial intelligence, vol 2475. Springer, Berlin, pp 331–338

Granular Neural Networks

Yan-Qing Zhang

Department of Computer Science, Georgia State University, Atlanta, GA, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Basic Architecture

Granular Learning Algorithms

Applications in Bioinformatics

Applications in Computational Web Intelligence

Applications in Brain Informatics

Conclusions

Future Directions

Bibliography

Glossary

Granular data Granular data include various data granules such as classes, clusters, subsets, groups, linguistic values, and intervals.

Granular neuron An artificial neuron maps granular data inputs to granular data outputs.

Granular link A granular link connects two different granular neurons.

Granular weights A granular weight represents connection strength between two granular neurons by using a granular value that is not limited to a traditional numerical value.

Granular neural network An intelligent neural network consists of granular neurons and granular links that connect relevant granular neurons.

Deep granular neural network A deep granular neural network consists of a large number of hidden layers with granular neurons and granular links that connect relevant granular neurons.

Definition of the Subject

Granular computing is a general computation theory for effectively using granules such as classes, clusters, subsets, groups, and intervals to build an efficient computational model for complex applications with huge amounts of data, information, and knowledge. Relevant existing techniques are divide and conquer, cluster analysis, fuzzy sets, rough sets, Yin-Yang computing, neutrosophic computing, interval computing, quotient space theory, belief functions, machine learning, databases, and so on.

A biological neural network in the human brain consists of a huge number of biological neurons that are networked to process various data sets (numerical and nonnumerical). Different complex data sets such as images, linguistic terms, numbers, patterns, and sounds can be recognized by biological neural networks. In other words, biological neural networks can process granules such as data clusters, linguistic words, information classes, and image groups from human sensor networks (i.e., eyes, ears, nose, etc.).

Based on granular computing and biological neural networks, a granular neural network is designed to deal with numerical-linguistic data fusion and granular knowledge discovery in numerical-linguistic databases. From a data granulation point of view, the granular neural network can process granular data in a database. From a data fusion point of view, the granular neural network makes decisions based on different kinds of granular data. From a knowledge discovery point of view, the granular neural network is able to learn internal granular relations between numerical-linguistic inputs and outputs and predict new relations in a database. The granular neural network is also capable of greatly compressing low-level granular data to high-level granular knowledge with some compression error and a data compression rate.

Introduction

Similar to a biological neural network in the human brain, a traditional artificial neural network consists of lots of artificial neurons that are connected by links. Each artificial neuron is simply a nonlinear mathematical mapping function, and each link has a numerical weight value showing connection strength between two neurons. The learning algorithm uses training data sets to optimize these weights of an artificial neural network, and the trained neural network can be used in prediction and decision-making applications. Traditional artificial neural networks based on the basic principle of biological neural networks in the human brain have been widely used in a lot of applications. However, traditional artificial neural networks use numerical values like 3.14 and -23.5 . Internal weights are also presented by numerical values that are not meaningful for users (i.e., the black box problem). Generally, the traditional artificial neural network has two long-term problems: (1) the black box problem: Learned knowledge in a neural network is represented by non-meaningful numerical weights (i.e., people cannot understand these numerical values because they are not linguistic values; therefore, people cannot understand how the neural network makes a decision), and (2) the curse of dimensionality: When the number of inputs and outputs is increased, the number of weights is increased dramatically (even exponentially). So the traditional artificial neural network cannot be used for large-scale applications with a large number of inputs and outputs.

Granular neural networks are used to process linguistic data like fuzzy terms to do granular knowledge discovery (Zhang et al. 2000). Granular neural networks with different structures can process information granules such as fuzzy sets and rough sets by using training algorithms (Pedrycz and Vukovich 2001). The granular neural network uses linguistic weights and the rules using linguistic arithmetic to speed up learning process (Dick and Kandel 2001). Granular neural networks are used to real applications such as land use classification (Vasilakos and Stathakis 2005). In summary, the granular neural learning algorithm can discover meaningful granular rules that can be understood

by ordinary people. So the granular neural network is useful to tackle the black box problem.

A set of training data is given to a computational model, and then the problem is how the network gleans useful knowledge from these training data sets in order to use that discovered knowledge to make right decisions for real applications. The two challenging problems for both computer science and cognitive science are the black box problem (i.e., the uninterpretability of a network of numerical weights or connection strengths) and the curse of dimensionality (i.e., the intractability complex cognitive problems for linear neural networks). This proposal plans to design a new architecture of a granular neural network with a new learning algorithm and then use data from both cognitive science and computer science to test and to improve the new learning algorithm in terms of the two challenging problems.

In recent years, hybrid neural networks such as fuzzy neural network, granular neural networks, and genetic neural networks have been investigated to solve complex application problems with high dimensionality. However, a single-stage hybrid neural network suffers from curse of dimensionality. A single-stage hybrid neural network suffers from curse of dimensionality because the number of parameters of the single-stage hybrid neural network is increasing exponentially with the increasing of the number of inputs. The hierarchical network structures based on multistage fuzzy reasoning have been used to solve the dimensionality problem. The ANFIS (Adaptive Network-Based Inference System), based on TSK fuzzy reasoning, is useful in many applications. Because it still cannot extract the commonly used fuzzy rules with both fuzzy IF part and fuzzy THEN part, it makes some difficulties in acquiring fuzzy knowledge from experts in a natural manner and learning commonly used fuzzy rules from data.

Therefore, how to design a powerful granular neural network with high learning speed, low training error, and low prediction error is still a challenging problem of artificial intelligence, computational intelligence, and cognitive science.

Basic Architecture

There are different types of granular neural networks. Now two granular neural networks are introduced in the two following sections, respectively.

First Architecture of Granular Neural Network

A database may contain numerical values, linguistic words, images, sounds, music pieces, and texts. The lowest data granulation technique deals with raw multimedia data collected directly from a real-world environment, whereas a higher data granulation technique classifies raw multimedia data into higher-level granules (i.e., classes, clusters, categories, groups, sets, etc.) to simplify data processing and data mining. For high-level data granulation, basic processing elements are these granules (not low-level raw data) (Zhang et al. 2000). In this sense, data granulation is related to data mining, data fusion, information fusion, and knowledge discovery. Interestingly, the human brain has a very strong ability to process multimedia granules, compute with linguistic words, and discover useful information from multimedia data. There should be links between raw multimedia data in an outside real world and biological neural networks inside a human brain. Similarly, links should be established between multimedia granules in databases and inputs-outputs of artificial neural networks. Since an artificial neural network cannot directly process these multimedia granules in most cases (because it is designed for directly processing numerical data), a conversion or interpretation of a granule, a link between a multimedia database and a neural network, becomes crucial to designing a granular neural network. In other words, how to convert multimedia granules into corresponding numerical features is very important for granular neural networks. For example, the linguistic data feature extraction system can convert fuzzy linguistic data such as very high, too small, and almost 100 into typical numerical features (Zhang et al. 2000).

Generally speaking, a granular neural network is capable of processing various granular data (granules). Granules could be a class of numbers,

a cluster of images, a set of concepts, a group of objects, a category of data, etc. These granules are inputs and outputs as multimedia data are inputs and outputs of biological neural networks in the human brain. Therefore, a granular-data-based granular neural network is more useful and more effective to process multimedia granules than conventional numerical-data-based neural network. Here, the granular neural network is made by Fuzzy Neural Networks with Knowledge Discovery (Zhang and Kandel 1998a).

Generally speaking, a granular neural network is capable of processing various granular data (granules) (Zhang et al. 2000). Granules could be a class of numbers, a cluster of images, a set of concepts, a group of objects, a category of data, etc. These granules are inputs and outputs of granular neural networks as multimedia data are inputs and outputs of biological neural networks in the human brain.

Therefore, granular-data-based granular neural networks are more useful and more effective to process multimedia granules than conventional numerical-data-based neural networks. Here, the FNN is a powerful Fuzzy Neural Network with Knowledge Discovery (FNNKD). The functions of different layers are described layer by layer as follows (Zhang et al. 2000, 2001):

Layer 1: Linguistic Feature Extraction Layer

In this layer, numerical-linguistic X and Y are transformed to corresponding fuzzy feature vectors $(a1, b1, c1, d1)$ and $(a2, b2, c2, d2)$, respectively.

Layer 2: Multi-FNNKD Layer

This layer consists of four dedicated FNNKDs (i.e., FNNKD1, FNNKD2, FNNKD3, and FNNKD4). FNNKD1 is a $2 \times k_1 \times 1$ fuzzy neural network which generates a crisp center a of an output fuzzy set by using k_1 fuzzy rules. FNNKD2 is a $2 \times k_2 \times 1$ fuzzy neural network which generates a crisp width b of an output fuzzy set by using k_2 fuzzy rules. FNNKD3 is a $2 \times k_3 \times 1$ fuzzy neural network which generates c of an output fuzzy set by using k_3 fuzzy rules. FNNKD4 is a $2 \times k_4 \times 1$ fuzzy neural network

which generates d of an output fuzzy set by using k_4 fuzzy rules.

An n -input-1-output normal fuzzy system has m fuzzy IF-THEN rules which are described by:

IF x_1 is A_1^k and $\dots$ and x_n is A_n^k THEN y is B^k ,
where x_i and y are input and output fuzzy linguistic variables, respectively.

Layer 3: Output Layer

Case 1: A fuzzy linguistic output is Z represented by a fuzzy feature vector (a, b, c, d) .

Case 2: A crisp numerical output is a .

For simplicity, the detailed learning algorithm is given in Zhang and Kandel (1998b). Once the learning procedure has been completed, all parameters for a FNNKD have been adjusted and optimized. As a result, all m fuzzy rules have been discovered from training data. Finally, the trained FNNKD can generate new values for new given input data.

Second Architecture of Granular Neural Network

Here, we plan to use traditional artificial intelligence; new computational intelligence including fuzzy logic, neural networks, evolutionary computation, and granular computing; and cognitive science to investigate the new evolutionary granular cognitive neural network that can discover hidden decision-making knowledge from data (i.e., know how the human brain learns new knowledge and makes a decision). The traditional neural learning algorithms can generate a lot of numerical weights that are not meaningful (i.e., the black box problem). Now we plan to design basic neuron granules that can contain meaningful knowledge like rules and make new granular learning methods based on cognitive science and computational intelligence (Lin et al. 2006).

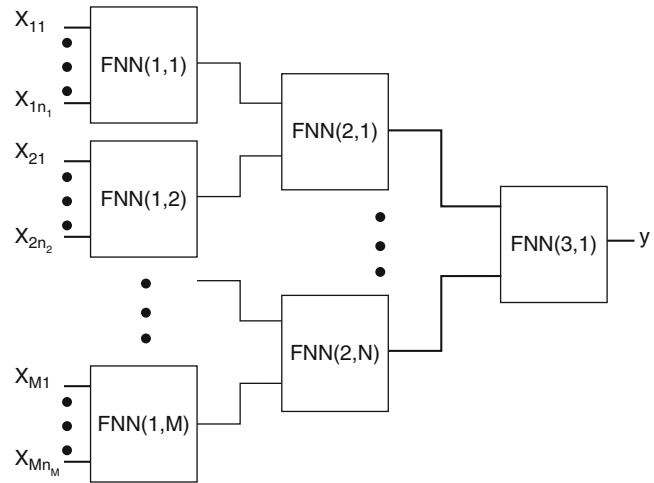
A new evolutionary granular cognitive neural network is proposed based on the normal fuzzy reasoning. The hybrid evolutionary granular cognitive neural learning algorithm using the divide-and-conquer strategy is developed to enhance learning quality in terms of discovered knowledge, training error, and prediction error.

Simulations have shown that the evolutionary granular cognitive neural network is an effective data mining and knowledge discovery system which discovers meaningful fuzzy rules, has low training error, and generates low prediction error. There are two main problems of curse of dimensionality for a soft computing system: (1) structural dimensionality (i.e., the total number of soft rules increases exponentially with the number of input variables in single-stage reasoning process) and (2) parametric dimensionality (i.e., the total number of adjustable system parameters increases exponentially with the number of input variables). For example, if a fuzzy system has n linguistic variables, each variable has m fuzzy linguistic sets, and if each fuzzy linguistic set has k parameters, then the fuzzy system will have m^n fuzzy rules and $(km)^n$ parameters. In this case, the structural complexity and the parametric complexity of the fuzzy system are $O(m^n)$ and $O[(km)^n]$, respectively. To solve the problems of curse of dimensionality, different hybrid soft computing architectures such as the multistage fuzzy neural network and the hierarchical system have been proposed to reduce system complexity in terms of dimensionality, time, and space. In general, the key technology for solving the curse of dimensionality of a hybrid soft computing system (particularly fuzzy neural network (FNN)) is the divide-and-conquer method. In other words, an original single-stage FNN with high system complexity can be divided into small multistage soft computing building blocks with low system complexity so as to reduce both structural dimensionality and parametric dimensionality significantly.

In general, the k th-stage FNN uses outputs generated from relevant $k - 1$ th-stage FNNs as inputs and then generates an output which will become an input of relevant $k + 1$ th-stage FNNs. A final stage FNN will generate a final output. For example, Fig. 1 shows the three-stage GCFNN. The FNN(i, j) is the i th-stage- j th FNN. So FNN(1, j) uses x_{1nj} as inputs for $j = 1, 2, \dots, M$, similarly FNN(2, l) for $l = 1, 2, \dots, N$ uses outputs generated from relevant first-stage FNNs as inputs, and finally FNN(3,1) generates the final output y . Therefore, the final output y is a function of all original inputs $x_{j1}, x_{j2}, \dots, x_{jnj}$ for $j = 1, 2, \dots, M$.

Granular Neural Networks,

Fig. 1 Architecture of a three-stage granular neural network (Zhang and Chung 2002)



Recently, hybrid neural networks such as fuzzy neural networks and granular neural networks have been investigated to solve complex application problems with high dimensionality. Such hybrid neural networks provide remarkable abilities to derive knowledge from complicated or imprecise data sets. Therefore, they have already been applied to extract patterns and detect trends that are too complex to be noticed by either human beings or other computer techniques. In this case, the trained neural networks can be thought of as an “expert” in the category of information it has been given to analyze.

In general, the key technology for solving the curse of dimensionality (including structural dimensionality and parametric dimensionality) of a hybrid soft computing system, such as a fuzzy neural network (FNN), is the divide-and-conquer method. In other words, an original single-stage FNN with high system complexity can be divided into small multistage soft computing building blocks with low system complexity so as to reduce both structural dimensionality and parametric dimensionality significantly.

In addition, GGCFNN provides parameters for deciding the number of iterations and termination criteria in order to control its learning performance. The hybrid genetic forward-wave-backward-wave learning algorithm is developed to enhance the system’s learning quality. Our simulation results generated by genetic granular cognitive fuzzy neural network are compared to

human persons and are analyzed in terms of computer science and cognitive science.

Deep Granular Neural Networks

Deep learning technology makes an artificial neural network with a large number of processing layers outperforms a conventional artificial neural network with few layers for various applications such as speech recognition, visual object recognition, drug discovery, and genomics (LeCun et al. 2015). Although a “deep granular neural network” is not a commonly used term in publications and media, basic properties of the granular neural network are used to build current deep neural networks such as a convolutional neural network that uses sub-images as granular data for inputs and outputs of different layers (LeCun et al. 1998). In recent years, convolutional neural networks have been successful for various important real-world applications such as deep neural networks that made more accurate skin cancer detection than dermatologists (Esteva et al. 2017), AlphaGO that beat top professional GO players (Silver et al. 2016), and deep convolutional neural networks that outperformed traditional machine learning models for the ImageNet competition in ILSVRC2012 (Krizhevsky et al. 2012). In 2014, *GoogLeNet* was the first place model that achieved 6.67% error rate for *ImageNet* Localization of ILSVRC2014 (Szegedy et al. 2015). In 2015, *ResNets* won the first place with 3.57% error rate for *ImageNet* Localization of ILSVRC2015

(He et al. 2016). In general, granules of images are transformed layer by layer through a large number of convolutional layers to greatly improve performance such as image classification accuracy.

Previously, different granular neural networks such as the granular neural network shown in Fig. 1 are described. If they use a large number of granular computational layers, they become deep granular neural networks. The new problem is how to make effective and efficient learning algorithms for the deep granular neural networks. For instance, more basic building blocks (FNNs) can be added in the three-layer granular neural network in Fig. 1 to make a deep granular neural network with a large number of layers. A new granular learning algorithm for training the deep granular neural network is shown in section “Granular Learning Algorithms.”

Granular Learning Algorithms

The traditional neural learning algorithms can generate a lot of numerical weights that are not meaningful (i.e., the black box problem). We plan to use traditional artificial intelligence, cognitive science, and new computational intelligence including fuzzy logic, neural networks, evolutionary computation, and granular computing to investigate the new evolutionary neural cognitive learning algorithms that can discover hidden decision-making knowledge from data (i.e., know how the human brain extracts rule-like relational knowledge from the inputs of experience and uses this knowledge in problem-solving and decision-making).

We plan to use traditional artificial intelligence, cognitive science, and computational intelligence to investigate the new granular cognitive neural network architecture with cognitive neural granules with a small number of meaningful linguistic parameters. Each cognitive neural granule consists of a number of artificial cognitive neurons with nonlinear numerical/linguistic mapping functions. So cognitive neural granules are basic

building blocks for the new granular cognitive neural network to effectively solve the problem of curse of dimensionality. The effective topological structure of the granular cognitive neural network will be investigated.

Suppose that the GCFNN has K inputs and one output. Given input data vectors X^p (i.e., $X^p = (X_1^p, X_2^p, \dots, X_K^p)$) and an output data vector Y^p for $p = 1, 2, \dots, L$, the global energy function is defined by

$$GE^p = \frac{1}{2} [f(X_1^p, \dots, X_K^p) - Y^p]^2. \quad (1)$$

The basic learning algorithm for the GCFNN was described in Zhang and Kandel (1998b). Here, for clarity and convenience, the relevant algorithms are introduced below. The new hybrid genetic-algorithms-based learning algorithm will be proposed in section “Forward-Wave-Backward-Wave Learning.”

Local Forward-Wave Learning

In general, suppose that a FNN has n inputs and one output. Given input data vectors x^p (i.e., $x^p = (x_1^p, x_2^p, \dots, x_n^p)$) and one-dimensional output data vector y^p for $p = 1, 2, \dots, N$, the energy function is defined by

$$E^p = \frac{1}{2} [f(x_1^p, \dots, x_n^p) - y^p]^2 \quad (2)$$

For simplicity, let E and f^p denote E^p and $f(x_1^p, x_2^p, \dots, x_n^p)$, respectively.

Based on the learning algorithm in Zhang and Chung (2002) and Zhang and Kandel (1998a), basic steps of the local forward-wave learning algorithm for a FNN are given below:

Procedure of Local Forward-Wave Learning

Step 1: Begin.

Step 2: Heuristic initialization of parameters.

Step 3: Gradient descending learning from data.

Then we can get the following learning algorithms

for $i = 1, 2, \dots, n, k = 1, 2, \dots, m, p = 1, 2, \dots, N, t = 0, 1, 2, \dots$, and a learning rate $\lambda > 0$.

Step 3.1: Train b^k

$$b^k(t+1) = b^k(t) - \lambda \frac{\partial E}{\partial b^k} | t, \quad (3)$$

Step 3.2: Train η^k

$$\eta^k(t+1) = \eta^k(t) - \lambda \frac{\partial E}{\partial \eta^k} | t, \quad (4)$$

Step 3.3: Train a_i^k

$$a_i^k(t+1) = a_i^k(t) - \lambda \frac{\partial E}{\partial a_i^k} | t, \quad (5)$$

Step 3.4: Train σ_i^k

$$\sigma_i^k(t+1) = \sigma_i^k(t) - \lambda \frac{\partial E}{\partial \sigma_i^k} | t, \quad (6)$$

Step 3.5: Train $wleft_i^k$ or $wright_i^k$

$$\begin{aligned} \text{IF } x_i^k \leq a_i^k \\ \text{THEN } wleft_i^k(t+1) = wleft_i^k(t) - \lambda \frac{\partial E}{\partial wleft_i^k} | t, \end{aligned} \quad (7)$$

$$\text{ELSE } wright_i^k(t+1) = wright_i^k(t) - \lambda \frac{\partial E}{\partial wright_i^k} | t. \quad (8)$$

Step 4: End.

Global Backward-Wave Learning

Based on the traditional back-propagation learning method, the global backward-wave learning algorithm is proposed to train all local FNNs in the back-propagation manner. The following procedure is a general algorithm for updating parameters of any FNN in GCFNN.

Procedure of Global Backward-Wave Learning

Step 1: Begin.

Step 2: Calculate a back-propagation error δ^l to the relevant $(l-1)$ th stage FNNs based on an error $\delta^{(l+1)}$ from the $(l+1)$ th stage FNN.

$$\delta^l = \frac{\partial E}{\partial y^{l-1}} = \delta^{(l+1)} \frac{\partial h^l}{\partial y^{l-1}}, \quad (9)$$

where y^l and $y^{(l-1)}$ are outputs of the l th stage FNN and the $(l-1)$ th stage FNN, respectively.

Step 3: Update parameters using the back-propagation error δ^l .

Step 3.5: Train $wleft_i^k$ or $wright_i^k$.

Step 4: Discover fuzzy knowledge.

Step 5: End.

Forward-Wave-Backward-Wave Learning

The new hybrid learning algorithm for training is called *forward-wave-backward-wave learning algorithm*, which combines the techniques of genetic algorithms, local forward-wave learning, and global backward-wave learning.

Procedure Genetic Forward-Wave-Backward-Wave Learning

Step 1: Begin.

Step 2: Partition the original training data sets input data vectors X^p and an output data vector Y^p for $p = 1, 2, \dots, L$ into M local training data sets where M is the number of first-stage granular neural networks.

Step 3: Local forward-wave learning: *local-forward-wave-learning()*.

Step 4: Global backward-wave learning: *global-backward-wave-learning()*.

Step 5: End.

Applications in Bioinformatics

A new granular neural network is used to form the new tertiary architecture (Reyaz-Ahmed and Zhang 2007). The neurons in the new granular neural network are SVM machines that classify the input data into two classes of protein

structures. The two classes are the binary classes that the SVM machines are actually trained for. The tertiary classifier's architecture is explained in subsequent sections.

The new tertiary classifier makes use of both one-versus-one and one-versus-rest binary classifiers. The novel architecture makes use of all the six binary classifiers in neural net architecture. The architecture is shown in Fig. 2 (Reyaz-Ahmed and Zhang 2007).

There are two hidden layers. The output of the first one is the same as the output of the individual SVM.

Outputs of first hidden layer:

$$O_1 = SVM(H/\sim H)$$

$$O_2 = SVM(E/\sim E)$$

$$O_3 = SVM(C/\sim C).$$

The output of the second hidden layer considers the output of the first layer as well as the SVM machine stored in that layer. For example, the output of neuron 4 has an SVM binary classifier that positively classifies H and negatively classifies E ; the result of this SVM is combined with that of the first layer outputs. This method uses the outputs of the three one-versus-rest classifier in a single neuron.

In the formulations, the output of the second hidden layer is formed by adding the output of the SVM sitting inside the neuron with the product of

the weight and output of the corresponding neuron (i.e. the neuron which positively classifies the same class as the current neuron) and by subtracting the products of the other two neurons in the first layer.

The output of the second layer are calculated as

$$O_4 = SVM(H/E) + W_{41}O_1 - W_{42}O_2 - W_{43}O_3$$

In the above formula, we add the values of the SVM that positively classify the same class (H) and subtract those that positively classify other classes. Here W_{41} means weight between neuron 1 and neuron 4. Similarly, other weight corresponds to output, input naming pattern.

Similarly outputs of other two neurons in the second hidden layer are calculated as

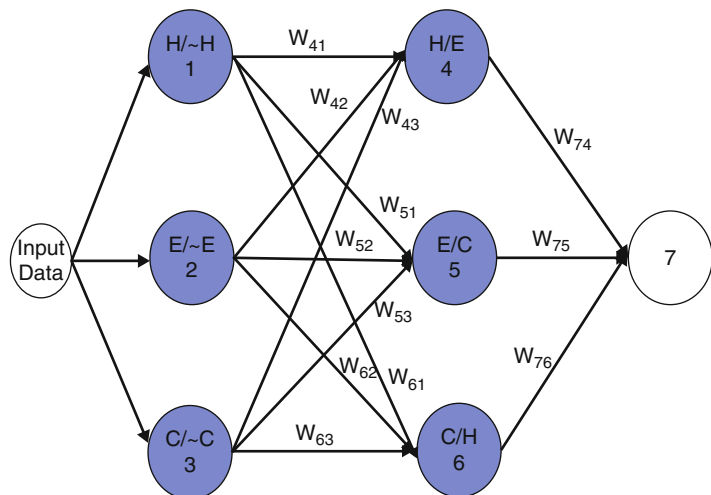
$$O_5 = SVM(E/C) + W_{52}O_2 - W_{51}O_1 - W_{53}O_3$$

$$O_6 = SVM(H/E) + W_{63}O_3 - W_{62}O_2 - W_{61}O_1$$

The final output layer does not have any SVM embedded in it. It calculates its results based on maximum of the three outputs of second hidden layer. There is only one neuron in this layer. The final output is one among the three classes (H , E , or C); whichever neuron produces the maximum output after multiplying it with appropriate weight with the second hidden layer output is considered as final output.

So the output of the third layer is as follows:

Granular Neural Networks, Fig. 2 Genetic neural support vector machines



If

$$(\text{Max}(W_{74}O_4, W_{75}O_5, W_{76}O_6) = W_{74}O_4)$$

Then

$$O_7 = H$$

Else If

$$(\text{Max}(W_{74}O_4, W_{75}O_5, W_{76}O_6) = W_{75}O_5)$$

Then

$$O_7 = E$$

Else

$$O_7 = C$$

For optimizing the weights, genetic algorithms are used. The weight range for the genetic granular neural network (GGNN) is selected to be between 0 and 1 so that the architecture performs to its full potential.

The same tertiary classifiers were demonstrated with binary classifiers using multiple windows scheme. This resulted in increase in total accuracy level, which is expected as the binary classifiers formed using multiple window scheme are better when compared to single window encoding scheme. The binary classifier used is constructed using three consecutive windows each of size 5 with gaps between the first and second window as well as between the second and third window. The results of these simulations are shown in Table 1.

Granular Neural Networks, Table 1 Accuracy comparisons

Tertiary classifier	Accuracy (%)
SVM_VOTE	62.6
SVM_REPRESNT.	64.8
GGNN	68.0

Applications in Computational Web Intelligence

Concepts of Input Data

Database tables have been generated using stock-historical data from www.yahoo.com site. Each line of the data set must contain five values date, open, high, low, close values of the stock. Each value is separated by space (see Table 2) (Zhang et al. 2001).

Overview of Implementation

A full run of the program implementation will be described, going through all the main features of the program (Zhang et al. 2001):

- Download historical data from the Internet.
- Copy the whole data into a text and run the program, which inserts the data into the database.
- A program is written through which users can buy and sell the stock shares, and the corresponding data is stored in the database.
- A program is written through which each user can see his/her own transactions.
- Algorithm, which trains the granular neural networks using the mean square error as stop criterion for learning, while never exceeding the maximum number of cycles which can take testing data from the initial date to the user entered date, and predicts the future stock closing values.
- A program is written, which compares the predicted values with real values.

One of the most important factors here is to construct a neural network deciding on what the

Granular Neural Networks, Table 2 Meaning of data parameters

dd/mm/yyyy	Date	Date of the day
Double	Open	Initial price
Double	Low	Lowest traded price
Double	High	Highest traded price
Double	Close	Price of the last trade
Double	Volume	Number of traded stocks

network will learn. A neural network must be trained on some input data. The two major problems in implementing the training are:

- Defining the set of input to be used (the learning environment)
- Deciding on an algorithm

Performance

The performance of the granular neural network algorithm is compared with the performance of the BP algorithm by training the same set of data and predicting the future stock values. If the training error was set at 0.03 and dow stock data were used by the BP algorithm and the granular learning algorithm, the granular neural network took 2 min and 58 s to train the neural network, whereas BP took 2 h and 55 min. The granular neural network's average error was 1.39, whereas BP gave 3.38.

The average error for granular neural network is less compared to the average error for BP algorithm. From the average error and the graph, it is conclusive that granular neural network produced closer future stock values with the real stock values compared to the BP algorithm using less training error. If the training error was set at 0.07 and dow stock data were used by the BP algorithm and the granular learning algorithm, the granular neural network took 2 min to train the neural network, whereas BP took 1 h and 48 min. The granular neural network's average error was 6.09, whereas BP gave 7.16.

The average error for the granular neural network is less compared to the average error for BP algorithm. From the average error and the graph, it is conclusive that the granular neural network produced closer future stock values with the real stock values compared to the BP algorithm using less training error. Based on the above two simulations, the overall performance with the granular neural network technique is better than BP technique.

Applications in Brain Informatics

Recent advances in granular computing, soft computing, and cognitive science have allowed an increase understanding of normal and abnormal brain functions, especially in the research of human's pattern recognition by means of computational intelligence.

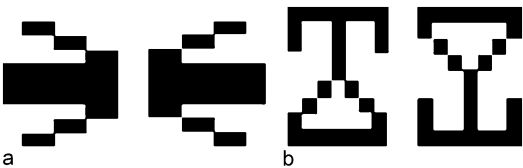
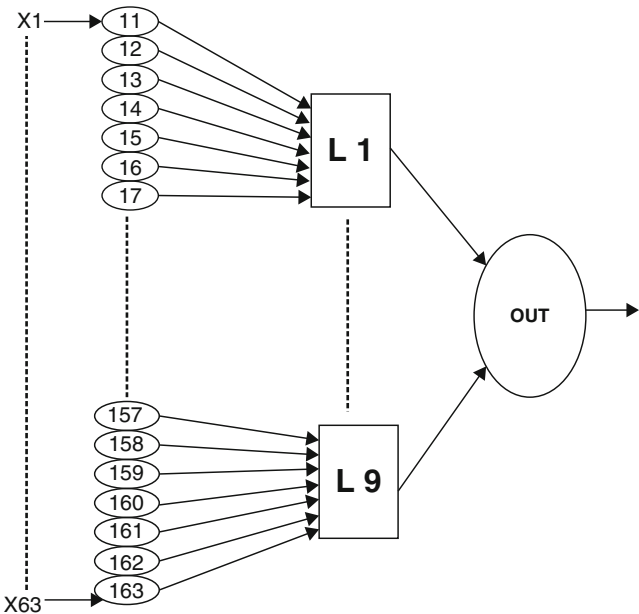
The simulations of the genetic granular neural network for pattern recognition have two ways: (1) testing whether a pair of noised pattern images can still be identified by the genetic granular neural network as similar as its original version after horizontal or vertical transformation and (2) testing whether the genetic granular neural network can recognize vertically or horizontally reversed pattern pairs with noise disturbance (Lin et al. 2006).

During each test, specific pattern samples are used for training and other groups of samples are served for testing. Each pattern sample is depicted by 7×9 pixel matrix, so totally 63 inputs are required in order to present one pattern, as shown in Fig. 3. To accommodate the input requirement, the genetic granular neural network's granular layer has 9 local genetic granular neural networks with 7 inputs, so that 63 pixel inputs are satisfied. In its second hidden layer, each genetic granular neural network has one output; the output layer has 9 inputs and 1 output. Therefore, the total number of inputs of the system is 126 (Lin et al. 2006).

Finally, the system will make a final decision whether or not two images in each testing sample are similar or symmetric based on the maximum output of the three genetic granular neural networks. Human beings are chosen to make their subjective judgment on pattern samples, by answering what degree the two patterns are similar or symmetric to each other (Lin et al. 2006).

Group A contains 50 pairs of arrow pattern samples within the following five noise levels, which are 0–10%, 10–20%, 20–30%, 30–40%, and 40–50% respectively. In each pair of arrow pattern sample, as shown in Fig. 4a, the second pattern is transformed from the first pattern after

Granular Neural Networks, Fig. 3 The genetic granular neural network structure for pattern recognition (Lin et al. 2006)



Granular Neural Networks, Fig. 4 (a) In Group A on the left, the second arrow pattern is transformed from the first after horizontal reversal; (b) In Group B on the right, the second cuplike pattern is transformed from the first after vertical reversal (Lin et al. 2006)

horizontal reversal at the first and then added into the same level of noises.

As shown in Fig. 4b, Group B includes 50 pairs of cuplike pattern samples within the same five noise levels as Group A. In each pair of cuplike pattern sample, the second pattern is transformed from the first pattern after vertical reversal and then added into the same level of noises.

Table 3 shows the comparisons of decisions made by human beings and the genetic granular neural network for scenario1. As can be seen, the genetic granular neural network and the human brain make 73% identical choices with a total 100 testing data. In Group A, only 10% of the 50 testing samples are different between the genetic granular neural network and human, while in Group B, almost half of the testing

Granular Neural Networks, Table 3 Simulation results

Group name	Different answers	Total samples	Agreement percentage
Group A	5	50	90%
Group B	22	50	56%
Total	27	100	73%

samples, which are more than four times higher than Group A, introduce contradictive decisions. Thus, human and computer systems in Group B make much more different answers than in Group A. It may indicate that the horizontal reversal causes more confusions than vertical reversal for human or the genetic granular neural network (Lin et al. 2006).

According to Tables 4 and 5, we discovered an interesting phenomenon: As noise level increases, the disagreement on pattern’s similarity between human and GGCNN tends to decrease. More specifically, as shown in Table 4, disagreement percentage is exactly 0% in (40%, 50%), while it increases to 16.67% in (0%, 30%). In Table 5, disagreement percentage is increased from 36% within the noise level (20%, 50%) to 55% in (0%, 20%). Additionally, the disagreement percentages experience prominent increases or decreases within low noise levels.

Granular Neural Networks, Table 4 Different answers to vertical reversal within five noise levels

Noise level	Different answers	Total samples	Disagreement percentage
40–50% noise	0	10	0%
30–40% noise	0	10	10%
20–30% noise	1	10	10%
10–20% noise	1	10	10%
0–10% noise	3	10	30%

Granular Neural Networks, Table 5 Different answers to horizontal reversal within five noise levels

Noise level	Different answers	Total samples	Disagreement percentage
40–50% noise	3	10	30%
30–40% noise	4	10	40%
20–30% noise	4	10	40%
10–20% noise	4	10	40%
0–10% noise	7	10	70%

The genetic granular neural network has exhibited good learning performance when stimulating human’s pattern recognition in terms of symmetry and similarity. It achieved higher agreement ratios when recognizing symmetrical patterns than similar patterns, and vertical transformation than horizontal transformation, even within high noise levels. The preliminary experiments indicated that the genetic granular neural network is able to simulate human’s recognition ability and adapt its learning experience to new patterns and then achieve quite similar results as human beings.

Conclusions

The human brain has high intelligence to recognize different geometrical patterns in terms of

similarity and symmetry, but a systematic framework of biological neural network has not been established. The granular neural network aims at simulating functionality of the human brain in terms of processing granular data such as linguistic terms, clusters, sets, and classes.

Since the human brain consists of biological neural networks that are the major components performing intelligent tasks, it is very important to develop advanced algorithms and mathematical models for the granular neural networks.

Future Directions

In the future, the challenging problem is how to design more powerful granular neural networks for more complex applications effectively and efficiently. A lot of advanced intelligent techniques such as type-1/type-2 fuzzy sets (Karnik et al. 1999; Zadeh 1979), rough sets (Lin 1997, 1999; Lin et al. 2002), granular computing (Lin et al. 2002; Pedrycz 2001), granular systems (Zhang 2005), granular support vector machines (Tang et al. 2005), granular kernel machines (Jin et al. 2007), and neurofuzzy systems (Jang 1993; Zhang and Chung 2002; Zhang and Kandel 1998b) can be merged to design a hybrid granular neural networks with various functions and outstanding performance.

Bibliography

- Dick S, Kandel A (2001) Granular weights in a neural network. In: IFSA world congress and 20th NAFIPS international conference, Vancouver, vol 3, July 2001, pp 1708–1713
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR 2016), pp 770–778
- Jang J-SR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23(3):665–685

- Jin B, Zhang Y-Q, Wang BH (2007) Granular kernel trees with parallel genetic algorithms for drug activity comparisons. *Int J Data Mining Bioinform* 1(3):270–285
- Karnik NN, Mendel JM, Liang Q (1999) Type-2 fuzzy logic systems. *IEEE Trans Fuzzy Syst* 7:643–658
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: *Advances in neural information processing systems* 25, Lake Tahoe, pp 1097–1105
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- LeCun Y, Bengio Y, Hinton GE (2015) Deep learning. *Nature* 521:436–444
- Lin TY (1997) Granular computing: from rough sets and neighborhood systems to information granulation and computing in words. In: *European congress on intelligent techniques and soft computing*, 8–12 Sept 1997, pp 1602–1606
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh L, Kacprzyk J (eds) *Computing with words in information/intelligent systems*. Physica, Heidelberg, pp 184–200
- Lin TY, Yao YY, Zadeh L (2002) Data mining, rough sets and granular computing. Physica, Heidelberg
- Lin C, Li J, Barrett N, Zhang Y-Q, Washburn DA (2006) Genetic granular cognitive fuzzy neural networks and human brains for pattern recognition. In: *The WICI international workshop on web intelligence (WI) meets brain informatics (BI) (WImBI 2006)*, Beijing, 15–16 Dec 2006
- Pedrycz W (2001) Granular computing: an emerging paradigm. Physica, Heidelberg
- Pedrycz W, Vukovich G (2001) Granular neural networks. *Neurocomputing* 36:205–224
- Pedrycz W, Kandel A, Zhang Y-Q (1997) Neurofuzzy systems. In: Dubois D, Prade H (eds) *Fuzzy systems: modeling and control*. Kluwer, Norwell, pp 311–380
- Reyaz-Ahmed A, Zhang Y-Q (2007) Protein secondary structure prediction using genetic neural support vector machines. In: *Proceedings of IEEE 7th international conference on BioInformatics and BioEngineering*, Boston, 14–17 Oct 2007, pp 1355–1359
- Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Triessche GVD, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529:484–503
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of 2015 IEEE conference on computer vision and pattern recognition (CVPR 2015)*, pp 1–9
- Tang YC, Jin B, Zhang Y-Q (2005) Granular support vector machines with association rules mining for protein homology prediction. *Artif Intell Med (Special Issue Comput Intell Tech Bioinform)* 35(1–2):121–134
- Vasilakos A, Stathakis D (2005) Granular neural networks for land use classification. *Soft Comput* 9:332–340
- Zadeh LA (1979) Fuzzy sets and information granulation. In: Gupta N, Ragade R, Yager R (eds) *Advances in fuzzy set theory and applications*. Elsevier, St. Louis, pp 3–18
- Zhang Y-Q (2005) Constructive granular systems with universal approximation and fast knowledge discovery. *IEEE Trans Fuzzy Syst* 13(1):48–57
- Zhang Y-Q, Chung F (2002) Fuzzy neural network tree with heuristic back-propagation learning. In: *Proceedings of IJCNN of world congress on computational intelligence 2002*, Honolulu, May 2002, pp 553–558
- Zhang Y-Q, Kandel A (1998a) Compensatory genetic fuzzy neural networks and their applications. In: *Series in machine perception artificial intelligence*, vol 30. World Scientific, Singapore
- Zhang Y-Q, Kandel A (1998b) Compensatory neurofuzzy systems with fast learning algorithms. *IEEE Trans Neural Netw* 9(1):83–105
- Zhang Y-Q, Fraser MD, Gagliano RA, Kandel A (2000) Granular neural networks for numerical-linguistic data fusion and knowledge discovery. *IEEE Trans Neural Netw* 11(3):658–667
- Zhang Y-Q, Akkaladevi S, Vachtsevanos G, Lin TY (2001) Fuzzy neural web agents for stock prediction. In: *Proceedings of FLINT2001*, UC Berkeley, 14–18 Aug 2001, pp 101–105

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems

Lech Polkowski

Mathematics and Computer Science, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction: Basic Ideas

Similarity

Information and Decision Systems

Granulation of Knowledge: General Ideas and Technique of Rough Inclusions

Granulation by Means of Rough Inclusions

Calculus on Granules

Some Principal Fields of Applications

Complexity Issues

Future Directions

Bibliography

Glossary

Knowledge The notion of knowledge can be explained by assuming the structure of the perceivable world as a system of states of things. These states are related among themselves by relations which in turn form a network of interdependent combinations. States and relations as well as dependencies among them are reflected in knowledge: things are reflected in objects or notions, relations in notions or concepts, and states of things in sentences. Sentences form knowledge, Bocheński (1954); for instance, in Artificial Intelligence, a judiciously chosen set of sentences forms a knowledge base for a logical agent, see Russell and Norvig (2003).

Knowledge Representation This is a chosen symbolic system (language) by means of which notions are encoded and reasoning is formalized.

Description Logic A common in Artificial Intelligence logical system used for knowledge representation, see Baader et al. (2004); the variant which is interesting for us is constructed in the attribute–value context in which objects are described in terms of chosen attributes (features) and their values. Primitive formulas of *descriptor logic* are *descriptors* of the form $(a = v)$ where a is an attribute and v is one of its values. Descriptors are interpreted in the universe U of objects: the meaning $[(a = v)]$ of the descriptor $(a = v)$ is defined as the set $\{u \in U : a(u) = v\}$. Descriptors are extended to formulas of description logic with the help of connectives of sentential calculus: when α, β are formulas then $\alpha \vee \beta, \alpha \wedge \beta, \neg \alpha, \alpha \Rightarrow \beta$ are formulas as well. The meaning of formulas is defined by recursion: $[\alpha \wedge \beta] = [\alpha] \cap [\beta], [\alpha \vee \beta] = [\alpha] \cup [\beta], [\neg \alpha] = U \setminus [\alpha]$. The meaning of implication $\alpha \Rightarrow \beta$ depends on the chosen interpretation of implication; in case when it is interpreted as in sentential calculus, that is, as $\neg \alpha \vee \beta$, the meaning is already defined by the given formulas for $\vee, \neg$.

Rough Sets: Knowledge as Classification Rough set theory proposed by Pawlak (1982), see Pawlak (1991), understands knowledge as classification: a set of equivalence relations each of which is understood as a certain classification of objects into its equivalence classes is given. Together, these relations constitute the knowledge base $\mathcal{R}$. Objects that belong in one class $[u]_R$ of a relation R in knowledge base are *R-indiscernible*. Objects that belong in the intersection $\bigcap_{R \in \mathcal{R}} [u]_R$ are *$\mathcal{R}$ -indiscernible*: they cannot be discerned by means of the available in $\mathcal{R}$ knowledge.

Indiscernibility relations are formally defined for sets of relations: given $\mathcal{T} \subseteq \mathcal{R}$, the $\mathcal{T}$ -indiscernibility relation $ind(\mathcal{T})$ is defined

as the intersection $\cap_{R \in \mathcal{T}} R$. Concepts are divided into two classes: $\mathcal{T}$ -exact concepts are sets of objects which can be expressed as unions of $\mathcal{T}$ -indiscernibility classes; other concepts are $\mathcal{T}$ -rough. For each concept X , there exist: the greatest exact concept $\mathcal{I}$ contained in x (the $\mathcal{T}$ -lower approximation to X) and the smallest exact concept $\mathcal{I}$ which contains x (the $\mathcal{T}$ -upper approximation to X). Rough concepts are perceived as uncertain sets sandwiched between a lower and an upper approximations.

Information Systems Information systems are commonly used for representing information about a given world fragment and they are pairs of the form (U, A) where U is a set of objects – representing things – and A is a set of attributes; each attribute a is modeled as a mapping $a: U \rightarrow V_a$ from the set of objects into the value set V_a . For an attribute a , the equivalence relation R_a is the set $\{(u, v) \in U \times U : a(u) = a(v)\}$ and the collection $\mathcal{R} = (R_a : a \in A)$ is the knowledge base. For each set $B \subseteq A$ of attributes, the B -indiscernibility relation $ind(B)$ is the set $\{(u, v) \in U \times U : a(u) = a(v) \text{ for each } a \in B\}$. Each object $u \in U$ can be described in two ways: first, by means of its indiscernibility class $[u]_A = \{v \in U : (u, v) \in ind(A)\}$; next, by means of its descriptor logic formula $\phi^A(u) : \bigwedge_{a \in A} (a = a(u))$; as $[\phi^A(u)] = [u]_A$, both descriptions are equivalent and u is described up to its equivalence class.

Decision Systems A particular form of an information system, a triple (U, A, d) in which d is the *decision*, the attribute not in A , that does express the evaluation of objects by an external oracle, an expert. Attributes in A are called *conditional* which does express the fact that they are set by us as conditions describing objects in our language. The principal problem related to decision systems is to find a description of dependence between conditional knowledge given by (U, A) and the expert knowledge $(U, \{d\})$.

Decision Rules A *decision rule* is a formula in descriptor language that does express a particular relation among conditional attributes in the attribute set A and the decision d , of the form: $\bigwedge_{a \in A} (a = v_a) \Rightarrow (d = v)$ with the

semantics defined in (Glossary: “Description Logic”). The formula is *true*, or *certain* in case $[\bigwedge_{a \in A} (a = v_a)] = \cap_{a \in A} [(a = v_a)] \subseteq [(d = v)]$. Otherwise, the formula is *partially true*, or *possible*. An object o which is matching the rule, that is, $a(o) = v_a$ for $a \in A$, can be classified to the class $[(d = v)]$; often a partial match based on a chosen similarity measure has to be performed.

Similarity Similarity relations are weaker forms of indiscernibility: indiscernible objects are similar but not necessarily *vice versa*; an example due to Henri Poincaré (1905) explains this best: assume that points x, y in the real line are similar whenever their distance is less than a fixed threshold $\delta : |x - y| < \delta$. Then certainly each x is similar to x , and if x is similar to y then y is similar to x , but from x similar to y and y similar to z it need not follow that x is similar to $z : |x - z|$ can be close to 2δ . Relations of this kind are reflexive and symmetric: they were called *tolerance relations* in Zeeman (1965). Tolerance relations in the framework of rough set theory information systems were considered in Nieminen (1988) and tolerance reducts and related notions in information systems were discussed in Polkowski et al. (1994). Some authors relax tolerance relations to similarity relations that need not be symmetric: for instance, rough inclusions Polkowski (2003) need not be symmetric.

In many applications, metrics are taken as similarity measures: a metric on a set U is a function $d : U \times U \rightarrow \mathbb{R}^+$ into non-negative reals such that (1) $d(x, y) = 0$ if and only if $x = y$. (2) $d(x, y) = d(y, x)$. (3) $d(x, y) \leq d(x, z) + d(z, y)$ for each x, y, z in X (the *triangle inequality*); similarity induced by a metric d is usually of the form $d(x, y) < \delta$ for a certain $\delta > 0$, Poincaré-style.

Rough Inclusions A ternary relation μ on a set $U \times U \times [0, 1]$ which satisfies conditions: (1) $\mu(x, x, 1)$. (2) $\mu(x, y, 1)$ is a binary partial order relation on the set X . (3) $\mu(x, y, 1)$ implies that for each object z in U : if $\mu(z, x, r)$ then $\mu(z, y, r)$. (4) $\mu(x, y, r)$ and $s < r$ imply that $\mu(x, y, s)$. The formula $\mu(x, y, r)$ is read as “the object x is a part in object y to a degree at least r .”

The partial containment idea does encompass the idea of an exact part, that is, mereological theory of concepts Leśniewski (1916). Similarity induced by a rough inclusion can be introduced as an asymmetric relation: $\mu(x, y, r)$, or a symmetric one: $\mu(x, y, r) \wedge \mu(y, x, r)$.

Granulation of Knowledge Idea of granulation was proposed by Zadeh (Zeeman 1965). In fuzzy calculus, granules are naturally induced as the inverse images of fuzzy membership functions and fuzzy computing is a fortiori computing with granules. A similar case happens with rough sets, as primitive objects in this paradigm are indiscernibility classes that are elementary granules, whereas their unions are granules of knowledge; reasoning in rough sets means by definition reasoning with granules. Each similarity relation induces granules understood as its classes; in comparison to indiscernibility relations, granules induced by similarity relations are more intricate due to lack of transitivity: relations among granules are more difficult to ascertain. In defining granules, attention is focused on the property that singles out objects belonging in the granule; usually, this property is being in the relation to an object chosen as the granule center: this means that distinct objects in the granule need not be in the relation. This feature distinguishes granules from *clusters*: aggregates in which any two objects have the defining property, for example, distance less than the chosen threshold. Also relations between pairs of distinct granules are difficult to be ascertained contrary to the case of, for example, clusters in which case the distance between distinct clusters is kept above a fixed threshold.

Classification of Objects Classification means the problem of assigning to each element in a set of objects (test sample) of a class (of a decision) to which the given element should belong; it is effected on the basis of knowledge induced from the given collection of examples (the training sample). To perform this task, objects are mapped usually onto vectors in a multidimensional real vector space (feature space). Paradigms invented to perform classification are many: from heuristics based on

cognitive networks like neural networks, through probabilistic tools like Bayesian networks to tools based on distance like k-nn methods and prototype methods, see Duda et al. (2001), Kloesgen and Zytkow (2002).

Fusion of Knowledge Fusion of knowledge means a combination of knowledge coming from few distinct sources, for example, in a mobile robot fusion of knowledge by means of a Kalman filter, see, for example, Russell and Norvig (2003) consists in producing an output – a robot's predicted position at the current moment, on the basis of inputs – a robot's position estimated as the last moment preceding the current moment, the current controls, and sensor readings in the current position.

Reasoning Processes of reasoning include an effort by means of which sentences are created; various forms of reasoning depend on the chosen system of notions, symbolic representation of notions, forms of manipulating symbols, see, for example, Bocheński (1954).

Mereology Mereology, see Leśniewski (1916), is a theory of concepts based on the notion of a part instead – as with naive set theory – on the notion of an element. The relation of being a part is irreflexive and transitive; the secondary notion of an ingredient means either being a part or being the whole object and it sets a partial order on objects. By means of the notion of an ingredient, the notion of a class operator is introduced in Leśniewski (1916) which converts ontological notions, for example, collections of objects into an individual object; it is used by us in granule formation.

Definition of the Subject and Its Importance

The creator of Fuzzy Set Theory Lotfi A. Zadeh, proposed to compute with granules in Zadeh (1979). The idea was natural as fuzzy reasonings are carried out in terms of fuzzy membership functions. A fuzzy membership function μ_X maps a universe U of objects into the interval $[0,1]$ and it represents the membership in a set

X as a membership to a degree. The value $\mu_X(x) = r$ is interpreted as the statement that the object x is an element of the set X to the degree of r . The mapping $U \rightarrow U/\mu_X$ which sends each object x to its fiber $\mu_X^{-1}(\mu_X(x))$ identifies into the granule $g_\mu(r)$ all objects that belong in X to the degree r . All fuzzy constructs are then expressed in terms of those granules. In this sense, fuzzy reasoning is in a natural way reasoning with granules. Granules in this reasoning are constructed in a uniform way, that is, all objects in a granule share the same property of external character: they belong to an "oracle" X to the same degree; changing X produces a variety of granules to reason with. The relation forming any granule is an equivalence R_X : the universe is decomposed into fibers of a fuzzy membership function μ_X in question, and the knowledge base is composed of all relations R_X for subsets $X \subseteq U$ of the universe of objects.

The same conclusion concerns rough set reasoning: elementary objects in reasoning are indiscernibility classes – elementary granules which are elements of a partition of the universe of objects by an indiscernibility relation $ind(B)$ (see Glossary: "Information Systems"). These granules are the smallest objects which can be described in terms of attributes and their values, that is, in description logic and they are used in forming description of objects, in building decision rules and then classifiers as well as control algorithms. Lin (1988, 1989) was the first to recognize topological character of granules and to form the basic notion of a neighborhood system as the meaning of the collection of granules on the universe of objects, see also Lin (1992, 1997, 1999, 2003, 2005); Lin (1989) recognized the import of tolerance relations by discussing tolerance induced neighborhoods.

In all hybrid approaches involving fuzzy or rough sets along with neural networks, genetic algorithms, etc., one is therefore bound to compute with granules; this fact testifies to the importance of granular structures.

In search of adequate similarity relations, various forms of granules were proposed and considered as well as experimentally verified as to their

effectiveness. In information systems, as granules were proposed *templates*, that is, generalized descriptors of the form $(a \in W_a)$ where $W_a \subseteq V_a$ with the meaning $[(a \in W_a)] = \{u \in U : a(u) \in W_a\}$ in Nguyen (2000); clearly, templates are aggregates in ontological sense of descriptors, that is, they form "big" granules. Their usage is motivated by their greater descriptive force than that of descriptors; a judicious choice of sets W_a should allow for constructing of a similarity relation that would reflect well the decision dependence on conditional attributes.

As means for granule construction, rough inclusions have been considered; a rough inclusion is (see Glossary: "Rough Inclusions") a generic term for a ternary predicate that formalizes the phrase: "the object x is a part of the object y to a degree r ," introduced in Polkowski and Skowron (1994, 1997). Granulation by means of rough inclusions has been studied, for example, in Polkowski (2003, 2004a, b, 2005a, b, 2006, 2007a, b). The idea of granule formation comes from mereology and it rests on using the class operator (see Glossary: "Mereology").

Granules formed by rough inclusions are used in models of fusion of knowledge, rough-neural computing, and in building many-valued logics reflecting the rough set ideology in reasoning.

Introduction: Basic Ideas

Granulation of knowledge can be considered from a few angles:

1. General purpose of granulation
2. Granules from binary relations
3. Granules in information systems from indiscernibility
4. Granules from generalized descriptors
5. Granules from rough inclusions: mereological approach

1. General Purpose of Granulation Granulation of knowledge comes into existence for a few reasons: the principal one is founded on the underlying assumption of basically all paradigms, viz., that reality exhibits a fundamental continuity,

that is, objects with identical descriptions in the given paradigm should exhibit the same properties with respect to classification or decision making. For instance, fuzzy set theory assumes that objects with identical membership descriptions should behave identically and rough set theory assumes that objects indiscernible with respect to a group of attributes should behave identically, in particular they should fall into the same decision class.

Thus, granulation is forced by assumptions of the respective paradigm and it is unavoidable once the paradigm is accepted and applied. Granules induced in the given paradigm by necessity form the first level of granulation.

In the search for similarities better with respect to applications like classification or decision making, more complex granules are constructed, for example, as unions of granules of the first level, or more complex functions of them, resulting, for example, in fusion of various granules from distinct sources.

Among granules of the first two levels, some kinds of them can be exhibited by means of various operators, for example, the class operator associated with a rough inclusion.

2. Granules from Binary Relations Granulation on the basis of binary general relations as well as specializations to, for example, tolerance relations, has been studied by Lin (1988, 1989, 1992, 1997, 1999, 2003, 2005, 2006) in particular as an important notion of a neighborhood systems; see also Yi Yu Yao (2000, 2005). This approach extends the approach based on indiscernibility as a special case. A general form of this approach according to Yao exploits the classical notion of Galois connection: two mappings $f: X \rightarrow Y, g: Y \rightarrow X$ form a Galois connection between ordered sets $(X, <)$ and $(Y, <)$ if and only if the equivalence $x < g(y) \Leftrightarrow f(x) < y$ holds. In an information system (U, A) , for a binary relation R on U , one considers the sets $xR = \{y \in U : xRy\}$ and $Rx = \{y \in U : yRx\}$, called, respectively, the *successor neighborhood* and the *predecessor neighborhood* of x . These sets are considered as granules formed by a specific relation of being affine to x in the sense of the relation R . Other forms of granulation can be obtained by

comparing objects with identical neighborhoods, for example, $x \equiv y \Leftrightarrow xR = yR$, etc. Saturation of sets of objects X, Y with respect to the relation R leads to sets X^*, Y^* such that $X^* R = Y^* \Leftrightarrow RY^* = X^*$, forming a Galois connection. This approach is closely related to the Formal Concept Analysis of Wille (1982).

3. Granules in Information Systems from Indiscernibility Granules based on indiscernibility in information/decision systems are constructed as indiscernibility classes: given an information system (U, A) (see Glossary: “Information Systems” and “Decision Systems”), and the collection $IND = \{ind(B) : B \subseteq A\}$ of indiscernibility relations, each elementary granule is of the form $[u]_B = \{u \in U : (u, v) \in ind(B)\}$ for some B . Among those granules, there are minimal ones: granules of the form $[u]_A$ induced from the set A of all attributes. Granules $[u]_A$ form the finest partition of the universe U ; given the class $[u]_B$ and the class $[u]_{A \setminus B}$ we have $[u]_A = [u]_B \cap [u]_{A \setminus B}$ hence $[u]_B = \cup_{v \in [u]_B \cap DIS_{A \setminus B}(u)} [u]_A$ where $v \in DIS_{A \setminus B}(u)$ if and only if there exists an attribute $a \in A \setminus B$ such that $a(u) \neq a(v)$. It is manifest that granules $[u]_B$ can be arranged into a tree with the root $[u]_\emptyset = U$ and leaves of the form $[u]_A$.

Granules based on indiscernibility form a complete Boolean algebra generated by atoms of the form $[u]_A$: unions of these atomic granules are closed on intersections and complements and these operations induce into granules the structure of a field of sets.

Atomic granules are made into some important unions by approximation operators: given a concept $X \subseteq U$, the lower approximation $\underline{B}X$ to X over the set B of attributes is defined as the union $\cup\{[u]_B : [u]_B \subseteq X\}$; the operator $L_B : Concepts \rightarrow Granules$ sending X to $\underline{B}X$ is monotone increasing and idempotent: $X \subseteq Y$ implies $\underline{B}X \subseteq \underline{B}Y$ and $L \circ L = L$. Similarly, the upper approximation $\overline{B}X = \cup\{[u]_B : [u]_B \cap X \neq \emptyset\}$ to X over B makes some elementary granules into the union; the operator U^B sending concepts into upper approximations is also monotone increasing and idempotent.

4. Granules from Generalized Descriptors Some authors have made use of generalized descriptors called *templates*, see Nguyen (2000); a template is a formula $T : (a \in W_a)$, where $W_a \subseteq V_a$ (see Glossary: “Information Systems”) with the meaning $g(T) : \{u \in U : a(u) \in W_a\}$. The granule $g(T)$ can be represented as the union $\cup_{u \in W_a} [u]_a$; granules of the form $[u]_a$ are also called *blocks*, see Grzymala-Busse (2004; Grzymala-Busse and Hu 2000).

5. Granules from Rough Inclusions: Mereological Approach Rough inclusions (see Glossary: “Rough Inclusions”) are ternary predicates of the form $\mu(x, y, r)$ which read: “ x is a part of y to degree at least r ”; intuitively, one can expect of μ the fulfillment of conditions:

- a. $\mu(x, x, 1)$
- b. $\mu(x, y, 1)$ if and only if $x \text{ ing}_\pi y$, where π is a part relation (see Glossary: “Mereology”) and ing_π is the associated ingredient relation: $x \text{ ing}_\pi y$ if and only if $x \pi y$ or $x = y$
- c. If $\mu(x, y, 1)$ then for each z : if $\mu(z, x, r)$ then $\mu(z, y, r)$ which is a condition of monotonicity of μ
- d. If $r > s$ and $\mu(z, y, r)$ then $\mu(z, y, s)$

Granule $g_r(u)$ of the radius r about the center u is defined as the class of property $\Phi(v) : \mu(v, u, r)$. The class operator $Cls : \text{Ontological entities (collections of objects)} \rightarrow \text{Mereological entities (individual objects)}$ acts on collections, or properties of objects and yields individual objects; an example of such class operator in set theory with the relation $\subset$ of part and the associated relation $\subseteq$ of ingredient is the union of a family of sets operator $\cup$: it converts a family (a property) of sets $\mathcal{F}$ into its union $\cup \mathcal{F}$ which is a single set.

The class operator Cls applied to a nonvacuous property Φ of objects yields the object $Cls \Phi$ which is defined as the unique object with the properties:

1. if $\Phi(u)$ then $\text{ing}_\pi Cls \Phi$

2. if $u \text{ ing}_\pi Cls \Phi$ then there exist objects v, w such that $v \text{ ing}_\pi u, v \text{ ing}_\pi w, \Phi(w)$

In other words, $Cls \Phi$ absorbs all objects whose each part has a part in common with an object which satisfies Φ .

Hence: $g_r(u)$ is $Cls \{u : \mu(v, u, r)\}$. Properties of granules defined in that way depend on properties of the rough inclusion μ .

There are many rough inclusions to select from, and basically they are defined from one of three sources, see Polkowski (2003, 2004a, b, 2005a, b, 2006, 2007a, b)

A From archimedean t-norms: an archimedean t-norm t see Polkowski (2002), admits a representation $t(x, y) = g_t(f_t(x) + f_t(y))$ with a continuous decreasing $f_t : [0, 1] \rightarrow [0, 1]$ and its pseudo-inverse g_t , see Ling (1965) or Polkowski (2002).

We define the rough inclusion μ_t as follows:

$$\mu_t(u, v, r) \text{ if and only if } g_t\left(\frac{|DIS_A(u, v)|}{|A|}\right) \geq r,$$

where $DIS_A(u, v) = \{a \in A : a(u) \neq a(v)\}$.

The most important example of rough inclusions obtained in this way is the rough inclusion μ_L induced from the Łukasiewicz t-norm $t_L(x, y) = \max\{0, x + y - 1\}$ for which $f(x) = 1 - x$ and $g(y) = 1 - y$, see Ling (1965) or Polkowski (2002). Introducing the complement $IND_A(u, v) = U \times U \setminus DIS_A(u, v)$, one can write the rough inclusion μ_L as,

$$\mu_L(u, v, r) \text{ if and only if } \frac{|IND_A(u, v)|}{|A|} \geq r$$

Thus, objects u, v are similar to degree r at least if and only if the probability of choosing randomly an attribute which does not discern between u and v is r at least. This rough inclusion is also induced by the relative Hamming distance on information sets of objects in the set U of the information system see Polkowski (2007a).

B From continuous t-norms see Polkowski (2007a): a continuous t-norm t induces the *residual implication* $\Rightarrow_t$ by the formula,

$x \Rightarrow_t y \leq z$ if and only if $t(x, z) \leq y$,

and it is well known that $x \Rightarrow_t y = 1$ if and only if $x \leq y$. Thus, residual implications are proper candidates for rough inclusions, once proper set functions $\Phi(u, v)$, $\Psi(u, v)$ are found such that $\mu_r(u, v, r)$ holds if and only if $\Phi(u, v) \Rightarrow_t \Psi(u, v) \geq r$ holds. Examples are given in Polkowski (2007a).

C It is evident that for a metric d on the universe U , the predicate $\mu_d(u, v, r)$ satisfied if and only if $d(u, v) \leq 1 - r$ is a rough inclusion. In particular, for the Hamming metric relative to the set A of attributes, $h(u, v) = \frac{|\{a \in A: a(u) \neq a(v)\}|}{|A|}$, the rough inclusion μ_h coincides with μ_L .

A rough inclusion μ_t is *transitive* in case the property holds,

If $\mu_t(u, v, r)$ and $\mu_t(v, w, s)$ then $\mu_t(u, w, t(r, s))$, and it was proved that rough inclusions obtained by methods 1 and 2 are transitive, Polkowski (2007a). One checks also that rough inclusions obtained from metrics by method 3 are transitive with the Łukasiewicz rough inclusion as the transitivity measure: for a metric d and $\mu_d(u, v, r)$, $\mu_d(v, w, s)$, it follows that $d(u, v) \leq 1 - r$ and $d(v, w, s) \leq 1 - s$ and the triangle inequality implies that $d(u, w) \leq (1 - r) + (1 - s)$ hence $\mu_d(u, w, r + s - 1)$, that is, $\mu_d(u, w, t_L(r, s))$.

Reasoning about granules defined from rough inclusions is carried out on lines of mereological deduction rule Leśniewski (1916):

(D(eduction) R(ule) in M(ereology)) For objects x, y : if for each object z , from $z \text{ ing}_\pi x$ it follows that there exists an object w such that $w \text{ ing}_\pi z$ and $z \text{ ing}_\pi y$ then $x \text{ ing}_\pi y$.

Similarity

Similarity relations came to attention of eminent theorists in the beginning of twentieth century due to interest in problems of mechanisms of thinking and perception: Henri Poincaré (1905) considered relations of discernibility, for example, two points in the space are *unintelligible* if and only if their distance is less than a fixed threshold value δ . This relation is certainly reflexive (a point cannot be discerned from itself) and symmetric, but may

lack the transitivity property; again, in connection with perception mechanisms, these relations were called *tolerance relations* in Zeeman (1965).

Formally, a *tolerance relation* τ is a relation which is reflexive: $\tau(x, x)$ for each x , and symmetric: if $\tau(x, y)$ then $\tau(y, x)$ for each pair x, y .

Tolerance relations are weaker than equivalence relations of, for example, indiscernibility (see Glossary: “Information Systems”); lack of transitivity makes them more difficult to analyze. In analogy to equivalence relations, one introduces *tolerance classes*: a tolerance class c_τ is defined as a maximal set with the property: for each pair $x, y \in c_\tau$ the relation $\tau(x, y)$ holds.

Contrary to equivalence classes, distinct tolerance classes can intersect: the collection of all pairwise distinct tolerance classes is a *covering* of the universe U of objects.

From the maximality principle in set theory (Teichmüller’s, Vaught’s or Zorn’s, see Kuratowski and Mostowski 1966), it follows that for each pair x, y with $\tau(x, y)$ there exists a tolerance class $c_{\tau(x, y)}$ contains both x, y .

We denote by the symbol $C_\tau(x)$ the collection of all tolerance classes which contain x . Then the principal property of tolerance relations holds: For each pair x, y of objects in the universe U and a tolerance relation τ on this universe: $\tau(x, y)$ if and only if $C_\tau(x) \cap C_\tau(y) \neq \emptyset$.

This fact shows that a canonical tolerance relation is defined as the nonempty intersection of sets, and each tolerance relation can be reduced to it.

Tolerance relations in information systems should factor through indiscernibility: if $\tau(u, v)$ and $u \text{ ind}(A) u', v \text{ ind}(A) v'$ then $\tau(u', v')$. Thus, tolerance relations in information systems should be defined in terms of attributes and their values.

Examples are: metrics-induced tolerance relations, for example, $\tau(u, v)$ if and only if $\max_a |a(u) - a(v)| < \delta$ for a fixed value of δ (tolerance based on the Manhattan metric) or $\tau(u, v)$ if and only if $\sum_a \left(|a(u) - a(v)|^2 \right)^{\frac{1}{2}} < \delta$ (tolerance based on the Euclidean metric).

Rough inclusions also supply examples of tolerance relations: given a rough inclusion μ , and a real number $r \in [0, 1]$, the tolerance relation $\tau_{\mu, r}$

can be defined as follows: $\tau_{\mu,r}(u,v)$ if and only if $\mu(u, v, r)$ and $\mu(v, u, r)$.

Yet weaker is the class of similarity relations which are merely reflexive: this kind of similarity relations encompass relations of rough inclusions, induced by asymmetric part relations as well as relations discussed in Słowiński et al. (Greco et al. 1999).

A reflexive similarity relation τ can be described in terms of classes $C_\tau(x) = \{y \in U : \tau(x, y)\}$ for $x \in U$. Then: $\tau(x, y)$ if and only if $y \in C_\tau(x)$. Thus, reflexive similarity relations can be canonically described as relations of membership between objects and their collections.

Information and Decision Systems

The scope of our discussion is limited to knowledge representation in the form of information systems (see Glossary: “Information Systems,” “Decision Systems,” “Description Logic,” and “Decision Rules”).

Information systems are means for formal rendering of data usually given in the form of data tables: rows in the table correspond to objects being described, and columns correspond to features used in object descriptions.

Each data table involves the set U of objects, called often the *universe of objects* and the set A of *attributes* representing formally features used in object descriptions. The pair (U, A) is called the *information system*.

For an object $u \in U$, and a set of attributes B , the *information set* of u is the set $inf_B(u) = \{(a = a(u)) : a \in B\}$.

Thus, the universe U can be represented as the set $Inf_B = \{inf_B(u) : u \in U\}$.

A principal assumption related to the information system (U, A) is that it supplies all information about objects and constructs derived from it should serve as evaluations of their extensions to the world of possible objects.

Therefore, objects u, v which satisfy the condition $inf_B(u) = inf_B(v)$ are regarded as *indiscernible* over B and perceived through the filter of attributes in B should be treated as identical with regard to description over B .

Formal rendering of this postulate is effected by means of *B-indiscernibility relation* $ind(B)$ (see Glossary: “Rough Sets: Knowledge as Classification”):

$$ind(B) = \{(u, v) \in U \times U : inf_B(u) = inf_B(v)\}.$$

It is an equivalence relation. Attribute sets produce a variety of indiscernibility relations with obvious properties like: $ind(C) \subseteq ind(D)$ if and only if $D \subseteq C$; $ind(C \cap D) = ind(C) \cap ind(D)$; $ind(A) \subseteq ind(C) \subseteq ind(\emptyset) = U \times U$ for each $C \subseteq A$.

Classes $[u]_B = \{u \in U : (u, v) \in ind(B)\}$ are minimal sets which can be described exactly over the set B : the class $[u]_B$ is the meaning of the descriptor formula $\bigwedge_{a \in B}(a = a(u))$.

Indiscernibility relations form the knowledge base from which inferences are derived about relations among objects, attributes, and attribute sets.

An important example of a property of attribute sets is the *reduct* property Pawlak (1985), Skowron and Rauszer (1992): a set B of attributes is a *reduct* if and only if B is a minimal with respect to set inclusion set of attributes such that $ind(B) = ind(A)$. Thus, each reduct does fully preserve classification and also knowledge.

The meaning of a reduct is also shown by means of dependencies among sets of attributes. The *functional dependency* between sets C and D of attributes, see Novotny and Pawlak (1988, 1992), Pawlak (1985), means that there exists a mapping $f_{C,D} : Inf_C \rightarrow Inf_D$ with the property that $f(inf_C(u)) = inf_D(u)$ for each $u \in D$. In this case, D *depends functionally* on C , in symbols, $C \mapsto D$.

Functional dependence $C \mapsto D$ means equivalently that $ind_C \subseteq ind_D$; in particular, as $ind_B \subseteq ind_C$ for each reduct B , the functional dependence $B \mapsto C$ takes place for each set C of attributes: each reduct determines functionally all values of attributes.

Finding reducts is computationally hard, see Skowron and Rauszer (1992), in theory reducts can be found as prime implicants of the Boolean discernibility function. Decision systems are (see Glossary: “Decision Systems”) information systems of the form $(U, A \cup \{d\})$ where $d \notin A$ is a *decision*.

The most important problem concerning decision systems consists in finding a set of *decision rules* (see Glossary: “Decision Rules”), which describes adequately the decision attribute in terms of attributes in A (*conditional attributes*).

Classification methods can be divided according to the adopted methodology, into classifiers based on reducts and decision rules, classifiers based on templates and similarity, classifiers based on descriptor search, classifiers based on granular descriptors, hybrid classifiers.

For a decision system (U, A, d) , classifiers are sets of decision rules. Induction of rules was a subject of research in rough set theory since its beginning. In most general terms, building a classifier consists in searching in the pool of descriptors for their conjuncts that describe sufficiently well decision classes. As distinguished in Stefanowski (2007), there are three main kinds of classifiers searched for: *minimal*, that is, consisting of minimum possible number of rules describing decision classes in the universe, *exhaustive*, that is, consisting of all possible rules, *satisfactory*, that is, containing rules tailored to a specific use. Classifiers are evaluated globally with respect to their ability to properly classify objects, usually by *error* which is the ratio of the number of correctly classified objects to the number of test objects, *total accuracy* being the ratio of the number of correctly classified cases to the number of recognized cases, and *total coverage*, that is, the ratio of the number of recognized test cases to the number of test cases.

Minimum size algorithms include LEM2 algorithm due to Grzymala-Busse (2004; Grzymala-Busse and Hu 2000) and covering algorithm in RSES package (Skowron et al. 1994); exhaustive algorithms include, for example, LERS system due to Grzymala-Busse (1992), systems based on discernibility matrices and Boolean reasoning by Skowron (1993), Bazan (1998), Bazan et al. (2000), implemented in the RSES package (Skowron et al. 1994).

Minimal consistent sets of rules were introduced in Skowron and Rauszer (1992); they were shown to coincide with rules induced on the basis of local reducts in Wróblewski (2004). Further developments include dynamic rules,

approximate rules, and relevant rules as described in Bazan (1998), Bazan et al. (2000) as well as local rules Bazan (1998), Bazan et al. (2000), effective in implementations of algorithms based on minimal consistent sets of rules. Rough set based classification algorithms, especially those implemented in the RSES system (Skowron et al. 1994), were discussed extensively in Bazan (1998); Skowron and Stepaniuk (2001) and Skowron and Swiniarski (2004) discuss rough set classifiers along with some attempt at analysis of granulation in the process of knowledge discovery.

In Bazan (1998), a number of techniques were verified in experiments with real data, based on various strategies:

discretization of attributes (codes: N-no discretization, S-standard discretization, D-cut selection by dynamic reducts, G-cut selection by generalized dynamic reducts)

dynamic selection of attributes (codes: N-no selection, D-selection by dynamic reducts, G-selection based on generalized dynamic reducts)

decision rule choice (codes: A-optimal decision rules, G-decision rules on basis of approximate reducts computed by Johnson’s algorithm, simulated annealing and Boltzmann machines etc., N-without computing of decision rules)

approximation of decision rules (codes: N-consistent decision rules, P-approximate rules obtained by descriptor dropping)

negotiations among rules (codes: S-based on strength, M-based on maximal strength, R-based on global strength, D-based on stability)

Any choice of a strategy in particular areas yields a compound strategy denoted with the alias being concatenation of symbols of strategies chosen in consecutive areas, for example, NNAND etc.

We record here in Table 1 an excerpt from the comparison (Tables 8, 9, and 10 in Bazan (1998)) of best of these strategies with results based on other paradigms in classification for two sets of

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 1 A comparison of errors in classification by rough set and other paradigms

Paradigm	System/method	Diabetes	Austr.credit
<i>Stat.Methods</i>	<i>Logdisc</i>	0.223	0.141
<i>Stat.Methods</i>	<i>SMART</i>	0.232	0.158
<i>Neural Nets</i>	<i>Backpropagation2</i>	0.248	0.154
<i>Neural Networks</i>	<i>RBF</i>	0.243	0.145
<i>Decision Trees</i>	<i>CART</i>	0.255	0.145
<i>Decision Trees</i>	<i>C4.5</i>	0.270	0.155
<i>Decision Trees</i>	<i>ITrule</i>	0.245	0.137
<i>Decision Rules</i>	<i>CN2</i>	0.289	0.204
<i>Rough Sets</i>	<i>NNANR</i>	0.335	0.140
<i>Rough Sets</i>	<i>DNANR</i>	0.280	0.165
<i>Rough Sets</i>	<i>Best result</i>	0.255(<i>DNAPM</i>)	0.130(<i>SNAPM</i>)

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 2 Accuracy of classification by template and similarity methods

Paradigm	System/method	Diabetes	Austr.credit
<i>Rough Sets</i>	<i>Simple.templ./Hamming</i>	0.6156	0.8217
<i>Rough Sets</i>	<i>Gen.templ./Hamming</i>	0.742	0.855
<i>Rough Sets</i>	<i>Simple.templ./Euclidean</i>	0.6312	0.8753
<i>Rough Sets</i>	<i>Gen.templ./Euclidean</i>	0.7006	0.8753
<i>Rough Sets</i>	<i>Match.tolerance</i>	0.757	0.8747
<i>Rough Sets</i>	<i>Clos.tolerance</i>	0.743	0.8246

data: Diabetes and Australian credit from UCI Repository (Skowron and Stepaniuk 2001).

An adaptive method of classifier construction was proposed in Wróblewski (2004); reducts are determined by means of a genetic algorithm and in turn reducts induce subtables of data regarded as classifying agents; choice of optimal ensembles of agents is done by a genetic algorithm.

Classifiers constructed by means of similarity relations are based on templates matching a given objects or closest to it with respect to a certain distance function, or on coverings of the universe of objects by tolerance classes and assigning the decision value on basis of some of them Nguyen S H (2000); we include in Table 2 excerpts from classification results in Nguyen S H (2000).

A combination of rough set methods with k-nearest neighbor idea is a further refinement of the classification based on similarity or analogy in Wojna (2005). In this approach, training set

objects are endowed with a metric, and the test objects are classified by voting of k nearest training objects for some k that is subject to optimization.

Granulation of Knowledge: General Ideas and Technique of Rough Inclusions

Granulation of knowledge means that given a representation of knowledge, for example, an information system, a fortiori, the set of information vectors of objects in its universe, these information vectors will undergo the process of aggregation, that is, of granule formation, granules being collections of information vectors. The process of granulation should be driven by a similarity relation on information vectors, and as the result of this process, similar to a satisfactory degree information vectors/objects will find

themselves in a one single granule. Parameters of this process are: an object u around which the granule is formed (the granule center) and the degree $r \in [0,1]$ to which objects in the granule are similar to u .

A General Form of Similarity

For our purposes of granulation of knowledge, we introduce a general form of graded similarity in the form a relation $\tau \subseteq U \times U \times [0, 1]$ on triples of the form (u, v, r) with $u, v \in U$, the universe of an information system (U, A) , and $r \in [0,1]$; we require of the similarity τ the following:

1. $\tau(u, u, 1)$
2. $r < s$ implies if $\tau(u, v, s)$ then $\tau(u, v, r)$
3. $\tau(u, v, 1)$ is a partial ordering of the universe U

Granulation: Tools

Given a similarity τ , we basically form a granule $g_\tau(u, r)$ of the radius r about an object u , by collecting together all objects v with the property that $\tau(v, u, r)$. Then, by definition of τ , one has the following properties of granules:

- a. $u \in g_\tau(u, r)$ for each $r \in [0, 1]$;
- b. $v \in g_\tau(u, r)$ implies $v \in g_\tau(u, s)$ for each $s < r$;

Granulation thus requires:

a similarity measure,
an aggregation mechanism.

Granulation by Means of Rough Inclusions

In this section, we apply the general ideas derived in sections “[Similarity](#)” and “[Information and Decision Systems](#)” in a specific context of rough inclusions (see Glossary: “[Rough Inclusions](#)”).

Mereological Theory of Concepts

As a first step to granulation by means of rough inclusions, we introduce the reader to mereological

theory of concepts which is the basis for our approach to granulation.

Mereology as conceived by Stanislaw Leśniewski (1916) is based on the primitive binary relation of *part*, π , on the universe U of objects, which is subject to the following requirements:

- (P1) $u \pi v$ implies $u \neq v$;
- (P2) $u \pi v$ and $v \pi w$ imply $u \pi w$.

Thus, the part relation π is nonreflexive and transitive.

Given a part relation, which as witnessed by (P1) means a proper part relation, one forms an improper part counterpart, called the ingredient, defined from the part relation as follows:

- (I) $v \text{ ing } u$ if and only if $v \pi u$ or $v = u$.

Thus, the ingredient relation is a partial ordering on the universe U :

- (U1) $u \text{ ing } u$ for each u
- (U2) $u \text{ ing } v$ and $v \text{ ing } u$ imply $v = u$
- (U3) $u \text{ ing } v$ and $v \text{ ing } w$ imply $u \text{ ing } w$

Mereology provides an operator, the *class operator*, whose purpose is to represent collections of objects as an object (individual); clearly, this will be also our principal granulation tool, when formal properties of granules are involved. The class operator Cls is applied to any nonempty collection (a property) Φ of individual objects in U , making this collection into an individual $Cls\Phi$; its definition is as follows Leśniewski (1916):

- (C11) If $v \in \Phi$ then $v \text{ ing } Cls \Phi$.
- (C12) If $v \text{ ing } Cls \Phi$ then for some $w, z \in U$ one has that $w \text{ ing } v, w \text{ ing } z, z \in \Phi$.

The object $Cls\Phi$ is unique for a given Φ .

As an example, the reader will verify that the relation $\subset$ of the proper set/concept containment is a part relation, the relation $\subseteq$ of the improper containment is the adjoint relation of an

ingredient, and given a nonempty family of sets Φ , the union $\cup\Phi$ is the class $Cls\Phi$ with respect to the ingredient relation $\subseteq$. Thus, the mereological context generalizes the standard set-theoretical context adopted prevalently by computer science.

Rough Inclusions

In the process of development of rough set theory, it has turned out that indiscernibility should be relaxed to similarity: in Polkowski et al. (1994), attention was focused on tolerance relations, that is, relations which are reflexive and symmetric but need not be transitive. An example of such relation was given in Poincaré (1905): given a metric ρ and a fixed small positive δ , one declares points x, y in the relation $sim(\delta)$ if and only if $\rho(x, y) < \delta$. The relation $sim(\delta)$ is a tolerance relation but it is equivalence for nonarchimedean ρ 's only, that is, when $\rho(x, y) \leq \max \{\rho(x, z), \rho(z, y)\}$.

We continue this example by introducing a graded version of $sim(\delta)$, viz., for a real number $r \in [0, 1]$, we define the relation $sim(\delta, r)$ by letting,

$$sim(\delta, r)(x, y) \text{ iff } \rho(x, y) \leq 1 - r. \quad (1)$$

The collection $sim(\delta, r)$ of relations have the following properties evident by the properties of the metric ρ :

- (SIM1) $sim(\delta, 1)(x, y)$ iff $x = y$;
- (SIM2) $sim(\delta, 1)(x, y)$ and $sim(\delta, r)(z, x)$ imply $sim(\delta, r)(z, y)$;
- (SIM3) $sim(\delta, r)(x, y)$ and $s < r$ imply $sim(\delta, s)(x, y)$.

Properties (SIM1)–(SIM3) induced by the metric ρ refer to the ingredient relation = whose corresponding relation of part is empty; a generalization can thus be obtained by replacing the identity with an ingredient relation *ing* in a mereological universe (U, π) .

In consequence a relation $\mu(u, v, r)$ is defined that satisfies the following conditions:

(RM1) $\mu(u, v, 1)$ iff $u \text{ ing } v$;

(RM2) $\mu(u, v, 1)$ and $\mu(w, u, r)$ imply $\mu(w, v, r)$;

(RM3) $\mu(u, v, r)$ and $s < r$ imply $\mu(u, v, s)$.

Any relation μ which satisfies the conditions (RM1)–(RM3) is called a *rough inclusion*. This relation is a similarity relation which is not necessarily symmetric, but it is reflexive. It is read as “the relation of a part to a degree.”

Rough Inclusions: Case of Information Systems The problem of methods by which rough inclusions could be introduced in information/decision systems has been studied in Polkowski (2003, 2004a, b, 2005a, b, 2006, 2007a, b). Here we recapitulate the results and add new ones. We recall that an *information system* is a method of representing knowledge about a certain phenomenon in the form of a table of data; formally, it is a pair (U, A) where U is a set of objects and A is a set of conditional attributes; any object $u \in U$ is described by means of its *information set* $Inf(u) = \{(a, a(u)) : a \in A\}$. The indiscernibility relation *Ind*, definable sets and nondefinable sets are defined from *Inf* as indicated in (Glossary: “Information Systems”).

Rough Inclusions from Metrics As observed in section “Introduction: Basic Ideas,” any metric ρ defines a rough inclusion μ_ρ by means of the equivalence $\mu_\rho(u, v, r) \Leftrightarrow \rho(u, v) \leq 1 - r$.

A very important example of a rough inclusion obtained on these lines is the rough inclusion μ_h with $h(u, v)$ being the reduced Hamming distance on information vectors of u and v , that is, $h(u, v) = \frac{|\{a \in A : (a, a(u)) \neq (a, a(v))\}|}{|A|}$, $|X|$ denoting the cardinality of the set X .

Thus, $\mu_h(u, v, r)$ if and only if $h(u, v) \leq 1 - r$; introducing sets $DIS(u, v) = \{a \in A : (a, a(u)) \neq (a, a(v))\}$ and $IND(u, v) = A \setminus DIS(u, v) = \{a \in A : a(u) = a(v)\}$, one can write down the formula for either μ_h as,

$$\mu_h(u, v, r) \Leftrightarrow \frac{|DIS(u, v)|}{|A|} \leq 1 - r, \quad (2)$$

or

$$\mu_h(u, v, r) \Leftrightarrow \frac{|IND(u, v)|}{|A|} \geq r. \quad (3)$$

The formula (3) witnesses that the rough inclusion μ_h is an extension of the indiscernibility relation Ind to a graded indiscernibility.

In a similar manner, one should be able to compute rough inclusions induced by other metrics standardly used on information sets like Euclidean, Manhattan, etc.

Rough inclusions induced by metrics possess an important property of *functional transitivity* expressed in general form by the rule,

$$\frac{\mu_\rho(u, v, r), \mu_\rho(v, w, s)}{\mu_\rho(u, w, L(r, s))}, \quad (4)$$

where $L(r, s) = \max \{0, r + s - 1\}$ is the Łukasiewicz t-norm, see Polkowski (Zadeh 1979). We offer a short proof of this fact: assuming that $\mu_\rho(u, v, r), \mu_\rho(v, w, s)$ which means in terms of the metric ρ that $\rho(u, v) \leq 1 - r, \rho(v, w) \leq 1 - s$; by the triangle inequality, $\rho(u, w) \leq (1 - r) + (1 - s)$, that is, $\mu_\rho(u, w, r + s - 1)$.

Rough Inclusions from Functors of Many-Valued Logics A functor $(t - \text{norm})t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is *nonarchimedean* in case the equality $t(x, x) = x$ holds for $x = 0, 1$ only; it is known (see e.g. Polkowski 2002) that only such t-norms are the Łukasiewicz L and the product t-norm $P(x, y) = x \cdot y$.

Each of these t-norms admits a functional representation: $t(x, y) = g(f(x) + f(y))$ (see, e.g., Polkowski 2002).

One defines a rough inclusion μ_t by letting, see Polkowski (Poincaré 1905; Polkowski 2003, 2004a, b, 2005a, b, 2006, 2007a),

$$\mu_t(u, v, r) \Leftrightarrow g\left(\frac{|DIS(u, v)|}{|A|}\right) \geq r. \quad (5)$$

In particular, in case of the t-norm L , one has $g(x) = 1 - x$ (see, e.g., Polkowski 2002), and thus the rough inclusion is expressed by means of the formula (3).

Other systematic method for defining rough inclusions is by means of residual implications of continuous t-norms, see Polkowski (2004a, b, 2005a, b, 2006, 2007a, b).

For a continuous t-norm t , the *residual implication* $x \Rightarrow_t y$ is a mapping from the square $[0, 1]^2$ into $[0, 1]$ defined as follows (see, e.g., Polkowski 2002),

$$x \Rightarrow_t y \geq z \text{ iff } t(x, z) \leq y; \quad (6)$$

thus, $x \Rightarrow_t y = \max \{z : t(x, z) \leq y\}$.

Proposition 1 The residual implication $x \Rightarrow_t y$ does induce a rough inclusion $\mu_t^\Rightarrow$ by means of the formula: $\mu_t^\Rightarrow(x, y, r)$ if and only if $x \Rightarrow_t y \geq r$ for every continuous t-norm t .

We include a short argument for the sake of completeness; clearly, $\mu_t^\Rightarrow(x, x, 1)$ holds as $x \Rightarrow_t y \geq 1$ is equivalent to $x \leq y$. Assuming $\mu_t^\Rightarrow(x, y, 1)$, that is, $x \leq y$, and $\mu_t^\Rightarrow(y, x, r)$, that is, $z \Rightarrow x \geq r$; hence, $t(z, r) \leq x$ we have $t(z, r) \leq y$, that is, $z \Rightarrow y \geq r$ so finally $\mu_t^\Rightarrow(x, y, r)$. Clearly, by definition, from $\mu_t^\Rightarrow(x, y, r)$ and $s < r$ it does follow that $\mu_t^\Rightarrow(x, y, s)$.

We recollect here the basic cases of rough inclusions obtained from most frequently applied t-norms. In all cases, $\mu_t^\Rightarrow(x, y, 1)$ if $x \leq y$ so the associated *ing* relation is $\leq$ and the underlying part relation is $<$. For $r < 1$, that is, $x > y$, one has

Case 1. $t = L$; in this case $x \Rightarrow_L y = \min \{1, 1 - x + y\}$, hence $\mu_t^\Rightarrow(x, y, r)$ if and only if $1 - x + y \geq r$.

Case 2. $t = P$ where $P(x, y) = x \cdot y$; in this case, $x \Rightarrow_P y = \frac{y}{x}$ when $x \neq 0$ and 1 when $x = 0$ hence $\mu_t^\Rightarrow(x, y, r)$ if and only if $y \geq x \cdot r$.

Case 3. $t = \min(x, y)$; in this case $x \Rightarrow_{\min} y$ hence $\mu_t^\Rightarrow(x, y, r)$ if and only if $y \geq r$.

Transitivity of Rough Inclusions It has been proved Polkowski (2003, 2004a, b, 2005a, b, 2006, 2007a, b) that all rough inclusions induced from either nonarchimedean or continuous t-norms t in the manner as above are transitive in the sense of the formula:

$$\frac{\mu(u, v, r), \mu(u, w, s)}{\mu(u, w, t(r, s))}.$$

Applications to Granulation of Knowledge Formal theory of rough inclusions allows for a formal mechanism of granulation of knowledge; we assume an information system (U, A) is given. Granulation of knowledge, proposed as a paradigm by Zadeh (1979), means grouping objects into collections called granules, objects within a granule being similar with respect to a chosen measure; granular computing means computing with granules in place of objects.

The mechanism of granule formation based on rough inclusions has been presented by the author in a few works, see, e.g. Polkowski (2003, 2004a, b, 2005a, b, 2006, 2007a, b) and we recall it here. We assume an information system (U, A) given. The basic tool in establishing properties of granules is the class operator of mereology, along with the Lesniewski deduction rule (DRM) of section “[Introduction: Basic Ideas](#).”

Given a rough inclusion μ on the universe U , for each object u and each $r \in [0, 1]$, the granule $g_\mu(u, r)$ of the radius r about u relative to μ is defined as the class of the property $\Phi(u, r, \mu) = \{v : \mu(u, v, r)\}$:

$$g_\mu(u, r) \text{ is } Cls\Phi(u, r, \mu). \quad (7)$$

In case of t-norm-induced rough inclusions, by their transitivity, the important property holds, see Polkowski (2007a):

$$v \text{ ing } g_\mu(u, r) \text{ iff } \mu(v, u, r), \quad (8)$$

that is, the granule $g_\mu(u, r)$ is $\{v : \Phi(u, v, \mu)(v)\}$ with no synergy effect.

Rough Inclusions on Granules Due to the feature of mereology that it operates (due to the class operator) only on level of individuals, one can extend rough inclusions from objects to granules; the formula for extending a rough inclusion μ to a rough inclusion $\bar{\mu}$ on granules is a modification of the mereological deduction rule of section “[Introduction: Basic Ideas](#)”:

(DRMG): $\bar{\mu}(g, h, r)$ if and only if for each object z with $z \text{ ing } g$ there exists an object w with $w \text{ ing } h$ such that $\mu(z, w, r)$.

The fact that $\bar{\mu}$ is a rough inclusion can be established by means of the rule (DRMG)

Modifications and Variants of Rough Inclusions In applications to be presented in works in this special session, some modified rough inclusions or weaker similarity measures will be instrumental, and we include a discussion of them here.

Modification by Means of Metrics on Attribute Values For the rough inclusion μ_L , the formula $\mu_L(v, u, r)$ means that $\frac{|IND(v, u)|}{|A|} \geq r$, that is, at least $r \cdot 100$ percent of attributes agree on u and v ; an extension of this rough inclusion depends on a chosen metric ρ bounded by 1 in the attribute value space V (we assume a simple case that ρ works for each attribute).

Then, given an $\varepsilon \in [0, 1]$, we let $\mu^\varepsilon(v, u, r)$ if and only if $|\{a \in A : \rho(a(v), a(u)) < \varepsilon\}| \geq r \cdot |A|$; it is manifest that μ^ε is a rough inclusion if ρ is a nonarchimedean metric, that is, $\rho(u, w) \leq \max\{\rho(u, v), \rho(v, w)\}$; otherwise the monotonicity condition rm^2 of section “[Information and Decision Systems](#)” need not be satisfied and this takes place with most popular metrics like Euclidean, Manhattan, etc.

In this case, a remedy is to define a rough inclusion μ^* as follows: $\mu^*(v, u, r)$ if and only if there exists an ε such that $\mu^\varepsilon(v, u, r)$. Then it is easy to check that μ^* is a rough inclusion.

Modification by Transfer Assume that $\Rightarrow_t$ is chosen, and for an information system (U, A) , with an ingredient relation ing on U , a mapping $\phi : U \rightarrow [0, 1]$ is given such that $\phi(u) \leq \phi(v)$ if and only if $u \text{ ing } v$. Then, the relation,

$$\mu_\phi(v, u, r) \text{ iff } \phi(u) \Rightarrow_t \phi(v) \geq r, \quad (9)$$

is a rough inclusion on U . We include a short proof of this fact: that $\mu_\phi(u, v, 1)$ is equivalent to $\phi(u) \leq \phi(v)$ hence to $u \text{ ing } v$ was observed in section “[Introduction: Basic Ideas](#),” Granules

from Rough Inclusions: Mereological Approach, B. Assuming that $\mu_\phi(u, v, 1)$ and $\mu_\phi(w, u, r)$, the proof that (RM2) (see section “[Granulation of Knowledge: General Ideas and Technique of Rough Inclusions](#)”) holds, that is, $\mu_\phi(w, v, r)$ goes like proof of (RM2) in Proposition 1 of section “[Granulation of Knowledge: General Ideas and Technique of Rough Inclusions](#).” Finally, the property (RM3) is evident.

As ϕ depends on one argument, let us select an object $s \in U$ and consider:

1. $\phi_1 = dis(u) = \frac{|\{a \in A: a(s) \neq a(u)\}|}{|A|}$; 2. $\phi_2 = ind(u) = \frac{|\{a \in A: a(s) = a(u)\}|}{|A|}$; 3. $\phi_3 = dis_\varepsilon(u) = \frac{|\{a \in A: \rho(a(s), a(u)) \geq \varepsilon\}|}{|A|}$; $\phi_4 = ind_\varepsilon(u) = \frac{|\{a \in A: \rho(a(s), a(u)) \leq \varepsilon\}|}{|A|}$, where ρ is a chosen metric on the set of attribute values V , and ε is a chosen threshold in $[0, 1]$.

Proposition 2 In all cases $i = 1, 2, 3, 4$, the relation $\mu_{\phi_i}(v, u, r)$ defined with ϕ_i as ϕ is a rough inclusion.

The reference object s can be chosen as an “ideal object,” that is, whose conditional class is contained in its decision class, etc. The formula (9) can be generalized by considering a set $S = \{s_1, \dots, s_k\}$ of reference objects.

Comparison of objects u, v on lines of this section need not lead to rough inclusions due to a possible violation of the property (RM2); yet, such variants are of importance as they allow for a direct comparison among objects, rules, and granules. We point to a generalization of rough inclusions introduced

We introduce for given objects u, v , and $\varepsilon \in [0, 1]$, factors: $dis_\varepsilon(u, v) = \frac{|\{a \in A: \rho(a(u), a(v)) \geq \varepsilon\}|}{|A|}$, and $ind_\varepsilon(u, v) = \frac{|\{a \in A: \rho(a(u), a(v)) < \varepsilon\}|}{|A|}$, where ρ is a metric on attribute value sets.

Then, we modify the formula (9) to the form,

$$v(u, v, r) \text{ iff } dis_\varepsilon(u, v) \rightarrow_t ind_\varepsilon(u, v) \geq r. \quad (10)$$

Clearly, v has properties: 1. $v(u, u, 1)$; 2. $v(u, v, r)$ and $s < r$ imply $v(u, v, s)$ but monotonicity property (RM2) need not hold. Rough inclusions and their weak variants will be essentially

exploited in data mining tasks presented in the sequel.

Calculus on Granules

We recall that we define a granule $g_\mu(u, r)$ about $u \in U$ of the radius r , relative to the rough inclusion μ , as follows:

$$g_\mu(u, r) \text{ is } Cls_\mu \prod^\mu(u, r), \quad (11)$$

where $\prod^\mu(u, r)$ is the property defined by means of,

$$\prod^\mu(u, r)(v) \text{ iff } \mu(v, u, r). \quad (12)$$

General properties of granules are collected below,

1. if $y \text{ ing } x$ then $y \text{ ing } g_{r,x}$;
2. if $y \text{ ing } g_{r,x}$ and $z \text{ ing } y$ then $z \text{ ing } g_{r,x}$;
3. if $\mu(y, x, r)$ then $y \text{ ing } g_{r,x}$;
4. if $s < r$ then $g_{r,x} \text{ ing } g_{s,x}$,

which follow from properties (RM1)–(RM3) of rough inclusions (see section “[Granulation of Knowledge: General Ideas and Technique of Rough Inclusions](#)”) and the fact that *ing* is a partial order, in particular it is transitive.

More generally,

$$\text{If } y \text{ ing } g_{r,x} \text{ then } g_{s,y} \text{ ing } g_{T(r,s)}x. \quad (14)$$

The last statement follows directly from the transitivity property of discussed rough inclusions and from the deduction rule (DRM) of section “[Introduction: Basic Ideas](#).”

It is natural to regard granule system $\{g_r^{\mu_i}(x) : x \in U; r \in (0, 1)\}$ as a neighborhood system for a topology on U that may be called the *granular topology*; we refer here to the idea of a neighborhood system introduced by Lin (1988, 1989).

In order to make this idea explicit, we define classes of the form $N^T(x, r) = Cls\left(\psi_{r,x}^{\mu_r}\right)$, where

$$\psi_{r,x}^{\mu_r}(y) \Leftrightarrow \exists s > r. \mu_T(y, x, s). \quad (15)$$

We declare the system $\{N^T(x, r) : x \in U : r \in (0, 1)\}$ to be a neighborhood basis for a topology θ_μ . This is justified by the following

Theorem 1 *Here are properties of the system $\{N^T(x, r) : x \in U : r \in (0, 1)\}$:*

1. $y \text{ ingr } N^t(x, r) \Rightarrow \exists \delta > 0. N^t(y, \delta) \text{ ingr } N(x, r)$;
2. $s > r \Rightarrow N^t(x, r) \text{ ingr } N^t(x, s)$;
3. $z \text{ ingr } N^t(x, r) \wedge z \text{ ingr } N^t(y, s) \Rightarrow \exists \delta > 0. N^t(z, \delta) \text{ ingr } N^t(x, r) \wedge N^t(z, \delta) \text{ ingr } N^t(y, s)$.

(16)

An argument for (16) is like follows.

For Property 1, $y \text{ ingr } N^t(x, r)$ implies that there exists an $s > r$ such that $\mu_t(y, x, s)$. Let $\delta < 1$ be such that $t(u, s) > r$ whenever $u > \delta$; δ exists by continuity of t and the identity $t(1, s) = s$. Thus, if $z \text{ ingr } N^t(y, \delta)$, then $\mu_t(z, y, \eta)$ with $\eta > \delta$ and $\mu_t(z, x, t(\eta, s))$ hence $z \text{ ingr } N^t(x, r)$.

Property 2 follows by (RM3) and Property 3 is a corollary to properties 1 and 2. This concludes the argument for (16).

Granule systems defined above form a basis for applications where approximate reasoning is a crucial ingredient.

We begin with a basic application in which approximate reasoning itself is codified as a many-world (intentional) logic where granules serve as possible worlds.

Some Principal Fields of Applications

Applications of the proposed mechanism of granulation have been indicated in many areas of reasoning under uncertainty, in particular in:

Reasoning about uncertainty in information/decision systems by means of rough mereological granular logics

Reasoning in distributed systems (fusion of knowledge)

Reasoning in cognitive schemes (rough neural hybrid systems)

Granular classifiers

We devote some space to these applications.

Reasoning About Uncertainty

The idea of a granular rough mereological logic, see Polkowski (2004a), Polkowski and Semeniuk-Polkowska (Polkowski and Skowron 1997), consists in measuring the meaning of a unary predicate in the model which is a universe of an information system against a granule defined to a certain degree by means of a rough inclusion. The result can be regarded as the degree of truth (the logical value) of the predicate with respect to the given granule. The obtained logics are intentional as they can be regarded as mappings from the set of granules (possible worlds) to the set of logical values in the interval $[0, 1]$, the value at a given granule regarded as the extension at that granule of the generally defined intension, see van Benthem (1988) for a general introduction to intentional logics.

For our purpose, it is essential to extend rough inclusions to sets; we use the t-norm L along with the representation $L(r, s) = g(f(r) + f(s))$ already mentioned. We denote these kind of inclusions with the generic symbol v .

For finite sets X, Y , we let,

$$v_L(X, Y, r) \text{ iff } g\left(\frac{|X \setminus Y|}{|X|}\right) \geq r; \quad (17)$$

as $g(x) = 1 - x$, see Polkowski (Duda et al. 2001), we have that $v_L(X, Y, r)$ holds if and only if $\frac{|X \cap Y|}{|X|} \geq r$. Let us observe that v_L is *regular*, that is, $v_L(X, Y, 1)$ if and only if $X \subseteq Y$ and $v_L(X, Y, r)$ only with $r = 0$ if and only if $X \cap Y = \emptyset$.

Thus, the ingredient relation associated with a regular rough inclusion is the improper

containment $\subseteq$, whereas the underlying part relation is the strict containment $\subset$.

Other rough inclusion on sets we exploit is the 3-valued rough inclusion v_3 defined via the formula, see Polkowski (2004a),

$$v_3(X, Y, r) \text{ iff } \begin{cases} X \subset Y \text{ and } r = 1 \\ X \cap Y = \emptyset \text{ and } r = 0 \\ r = \frac{1}{2} \text{ otherwise,} \end{cases} \quad (18)$$

We assume that an information/decision system (U, A, d) is given, along with a rough inclusion v on the subsets of the universe U ; for a collection of predicates (unary) Pr , interpreted in the universe U (meaning that for each predicate $\phi \in Pr$ the meaning $[\phi]$ is a subset of U), we define the intentional logic grm_v on Pr by assigning to each predicate ϕ in Pr its intension $I_v(\phi)$ defined by its extension $I_v^\vee(g)$ at particular granules g , as,

$$I_v^\vee(g)(\phi) \geq r \text{ iff } v(g, [\phi], r). \quad (19)$$

With respect to the rough inclusion v_L , the formula (19) becomes,

$$I_{v_L}^\vee(g)(\phi) \geq r \text{ iff } \frac{|g \cap [\phi]|}{|g|} \geq r. \quad (20)$$

The counterpart for v_3 is specified by definition (18).

We say that a formula ϕ interpreted in the universe U of an information system (U, A) is *true* at a granule g with respect to a rough inclusion v if and only if $I_v^\vee(g)(\phi) = 1$.

Thus, for every regular rough inclusion v , a formula ϕ interpreted in the universe U , with meaning $[\phi]$, is true at a granule g with respect to v if and only if $g \subseteq [\phi]$. In particular, for a decision rule $r: p \Rightarrow q$ in the descriptor logic, the rule r is true at a granule g with respect to a regular rough inclusion v if and only if $g \cap [p] \subseteq [q]$.

The formula $v(g, [\phi], r) = 1$ stating the truth of ϕ at g , v with v regular can be regarded as a condition of orthogonality type, with the usual consequences.

1. If ϕ is true at granules g, h then it is true at $g \cup h$.
2. If ϕ is true at granules g, h then it is true at $g \cap h$.
3. If ϕ, ψ are true at a granule g then $\phi \vee \psi$ is true at g .
4. If ϕ, ψ are true at a granule g then $\phi \wedge \psi$ is true at g .
5. If ψ is true at a granule g then $\phi \Rightarrow \psi$ is true at g for every formula ϕ .
6. If ϕ is true at a granule g then $\phi \Rightarrow \psi$ is true at g if and only if ψ is true at g .

The graded relaxation of truth is given obviously by the condition, a formula ϕ is *true to a degree at least r* at g , v if and only if $I_v^\vee(g)(\phi) \geq r$, that is, $v(g, [\phi], r)$ holds. In particular, ϕ is *false* at g , v if and only if $I_v^\vee(g)(\phi) \geq r$ implies $r = 0$, that is, $v(g, [\phi], r)$ implies $r = 0$. The properties 1-6 above do suggest that the notion of truth in rough mereological logics has with respect to connectives of logical calculi similar properties to those of classical sentential calculus. Therefore, we introduce the following semantics of sentential connectives $\neg, \vee, \wedge, \Rightarrow$.

1. $[\neg\alpha] = U \setminus [\alpha]$.
2. $[\alpha \vee \beta] = [\alpha] \cup [\beta]$.
3. $[\alpha \wedge \beta] = [\alpha] \cap [\beta]$.
4. $[\alpha \Rightarrow \beta] = (U \setminus [\alpha]) \cup [\beta]$.

With respect to this semantics, the following properties hold.

1. For each regular v , a formula α is true at g , v and only if $\neg\alpha$ is false at g , v .
2. For $v = v_L, v_3$, $I_v^\vee(g)(\neg\alpha) \geq r$ if and only if $I_v^\vee(g)(\alpha) \geq s$ implies $s \leq 1 - r$.
3. For $v = v_L, v_3$, the implication $\alpha \Rightarrow \beta$ is true at g if and only if $g \cap [\alpha] \subseteq [\beta]$ and $\alpha \Rightarrow \beta$ is false at g if and only if $g \subseteq [\alpha] \setminus [\beta]$.
4. For $v = v_L$, if $I_v^\vee(g)(\alpha \Rightarrow \beta) \geq r$ then $\Rightarrow_L(t, s) \geq r$ where $I_v^\vee(g)(\alpha) \geq t$ and $I_v^\vee(g)(\beta) \geq s$.

The functor $\Rightarrow_L$ in 4 is the Łukasiewicz implication of many-valued logic $\Rightarrow_L(t, s) = \min \{1, 1 - t + s\}$.

Further analysis should be split into the case of v_L and the case of v_3 as the two differ essentially with respect to the form of reasoning they imply.

Reasoning with v_L The last property 4 shows in principle that the value of $I_v^\vee(g)(\alpha \Rightarrow \beta)$ is bounded from above by the value of $\Rightarrow_L(I_v^\vee(g)(\alpha), I_v^\vee(g)(\beta))$.

This suggests that the idea of collapse due to S. Lesniewski can be applied to formulas of rough mereological logic in the following form: for a formula $q(x)$ we denote by the symbol q^* the formula q regarded as a sentential formula (i.e., with variable symbols removed) subject to relations $\neg q(x)^* \text{ is } (\neg q(x)^*)$ and $p(x) \Rightarrow q(x)^* \text{ is } p(x)^* \Rightarrow q(x)^*$. As the value $[q^*]_g$ of the formula $q(x)^*$ we admit the value of $\frac{|g \cap q(x)|}{|g|}$. Thus, the item 4 can be rewritten in the form.

$$I_v^\vee(g)(\alpha \Rightarrow \beta) \leq \Rightarrow_L([\alpha^*]_g, [\beta^*]_g). \quad (21)$$

The following statement is then obvious: if $\alpha \Rightarrow \beta$ is true at g then the collapsed formula has the value 1 of truth at the granule g . This gives a necessity condition for verification of implications of rough mereological logics: if $\Rightarrow_L([\alpha^*]_g, [\beta^*]_g) < 1$ then the implication $\alpha \Rightarrow \beta$ is not true at g . This concerns in particular decision rules: for a decision rule $p(v) \Rightarrow q(v)$, it follows that the decision is true on a granule g if and only if $[p^*]_g \leq [q^*]_g$.

Possibility and Necessity Possibility and necessity are introduced in rough set theory by means of approximations: the upper and the lower, respectively. A logical rendering of these modalities in rough mereological logics exploits the approximations. Denoting the lower approximation as $\underline{X} = \{u \in U : [u]_A \subseteq X\}$ and the upper approximation as $\overline{X} = \{u \in U : [u]_A \cap X \neq \emptyset\}$, we define two modal operators: M (possibility) and L (necessity) by means of their semantics.

To this end, we let

$$\begin{aligned} I_v^\vee(g)(M\alpha) &\geq r \text{ iff } v_L(g, [\alpha], r) \\ I_v^\vee(g)(L\alpha) &\geq r \text{ iff } v_L(g, [\alpha], r). \end{aligned} \quad (22)$$

Then we have the following criteria for necessarily or possibly true formulas.

A formula α is *necessarily true at a granule g* if and only if $g \subseteq [\alpha]$; α is *possibly true at g* if and only if $g \subseteq [\alpha]$.

This semantics of modal operators M, L can be applied to show that rough set structures carry the semantics of S5 modal logics, that is, the following relations hold at each granule g .

1. $L(\alpha \Rightarrow \beta) \Rightarrow [(L\alpha) \Rightarrow L(\beta)]$.
2. $L\alpha \Rightarrow \alpha$.
3. $L\alpha \Rightarrow LL\alpha$.
4. $M\alpha \Rightarrow LM\alpha$.

Proofs can be found in Polkowski (2004a).

Reasoning with v_3 In case of $v = v_3$, one can check on the basis of definitions that $I_v^\vee(g)(\neg\alpha) \geq r$ if and only if $I_v^\vee(g)(\alpha) \leq 1 - r$; thus the negation functor in rough mereological logic based on v_3 is the same as the negation functor in the 3-valued Łukasiewicz logic. For implication, the relations between rough mereological and 3-valued logics L_3 follow from tables of values of truth.

Table 3 shows truth values of implications for the 3-valued logic L_3 and Table 4 shows truth values of implications for rough mereological logic based on v_3 .

To express the relation between the two implications, we introduce a new notion: we say that a

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 3 Truth values for implication in L_3

$\Rightarrow$	0	1	$\frac{1}{2}$
0	1	1	1
1	0	1	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	1	1

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 4 Truth values for implication $\alpha \Rightarrow \beta$ in logic based on v_3

$\Rightarrow$	$I_v^\vee(g)(\beta) = 0$	$I_v^\vee(g)(\beta) = 1$	$I_v^\vee(g)(\beta) = \frac{1}{2}$
$I_v^\vee(g)(\alpha) = 0$	1	1	1
$I_v^\vee(g)(\alpha) = 1$	0	1	$\frac{1}{2}$
$I_v^\vee(g)(\alpha) = \frac{1}{2}$	$\frac{1}{2}$	1	1 when $g \cap [x] \subseteq [\beta]$; $\frac{1}{2}$ otherwise

formula ϕ is *acceptable* in either logic (L_3 or grm_{v_3}) at a granule g if and only if $[\phi^*]_g \geq \frac{1}{2}$, respectively, $I_v^\vee(g)(\phi) \geq \frac{1}{2}$.

From Tables 3 and 4, one infers that $I_v^\vee(g) \times (\phi) \geq [\phi^*]_g$. This crucial relationship implies that: if ϕ is acceptable at g then ϕ is acceptable at g ; if ϕ^* is true at g then ϕ is true at g . Also, if ϕ is false at g then ϕ^* is false at g .

This fact allows for a sufficiency criterion: if ϕ^* is a theorem of the logic L_3 then ϕ is a true formula of the logic grm_{v_3} . This fact supplies a characterization of decision rules as well.

Reasoning in Distributed Systems

We now begin with the first of principal applications of the developed so far apparatus, cf., Polkowski and Skowron (1999). Rough inclusions and granular intentional logics based on them can be applied in describing workings of a collection of intelligent agents which are called here granular agents as they are endowed with granular logics.

A granular agent ag in its simplest form is a tuple

$$ag^* = (U_{ag}, A_{ag}, \mu_{ag}, Pred_{ag}, UncProp_{ag}, GSynt_{ag}, LSynt_{ag}),$$

where,

1. $(U_{ag}, A_{ag}) = I_{ag}$ is an information system of the agent ag .
2. μ_{ag} is a rough inclusion induced from I_{ag} .
3. $Pred_{ag}$ is a set of first-order predicates interpreted in U_{ag} .
4. $UncProp_{ag}$ is the function that describes how uncertainty measured by rough inclusions at agents connected to ag propagates to ag .

5. The operator $GSynt_{ag}$, the granular synthesizer at ag , takes granules sent to the agent from agents connected to it, and makes those granules into a granule at ag ;
6. $LSynt_{ag}$, the logic synthesizer at ag , takes formulas sent to the agent ag by its connecting neighbors and makes them into a formula describing objects at ag .

A network of granular agents is a directed acyclic graph $N = (Ag, C)$, where Ag is its set of vertices, that is, granular agents, and C is the set of edges, that is, connections among agents, along with disjoint subsets $In, Out \subset Ag$ of, respectively, input and output agents.

We assume for simplicity that N consists of three agents connected into the tree, and we show a simple analysis of the direct fusion of knowledge; clearly, more complex schemes will require a deeper and more complex analysis but on the lines indicated in the example that follows.

Fusion of Knowledge: An Example We consider an agent $ag \in Ag$ and – for simplicity reasons – we assume that ag has two incoming connections from agents ag_1, ag_2 ; the number of outgoing connections is of no importance as ag sends along each of them the same information.

We assume that each agent is applying the rough inclusion μ_L induced by the Łukasiewicz t-norm L , see section “[Introduction: Basic Ideas](#),” Granules from Rough Inclusions: Mereological Approach, A. Each agent is also applying the rough inclusion on sets of the form (17) in evaluations related to extensions of formulae intensions.

Clearly, there exists a fusion operator o_{ag} that assembles from objects $x \in U_{ag1}, y \in U_{ag2}$ the

object $o(x, y) \in U_{ag}$; we assume that $o_{ag} = id_{ag1} \times id_{ag2}$, that is, $o_{ag}(x, y) = (x, y)$.

Similarly, we assume that the set of attributes at ag , equals: $A_{ag} = A_{ag1} \times A_{ag2}$, that is, attributes in A_{ag} are pairs (a_i, a_2) with $a_i \in A_{agi}(i = 1, 2)$ and that the value of this attribute is defined as:

$$(a_1, a_2)(x, y) = (a_1(x), a_2(y)).$$

It follows that the condition holds:

$$O_{ag}(x, y)IND_{ag}O_{ag}(x', y') \text{ iff } xIND_{ag1}x' \text{ and } yIND_{ag2}y'.$$

Concerning the function $UncProp_{ag}$, we consider objects x, x', y, y' ; clearly,

$$\begin{aligned} & DIS_{ag}(o_{ag}(x, y)o_{ag}(x', y')) \\ & \subseteq DIS_{ag1}(x, x') \times A_{ag2} \cup \\ & A_{ag1} \times DIS_{ag2}(y, y'), \end{aligned} \quad (23)$$

and hence,

$$\begin{aligned} & |DIS_{ag}(o_{ag}(x, y)o_{ag}(x', y'))| \\ & \leq |DIS_{ag1}(x, x') \cdot |A_{ag2}|| \\ & + |A_{ag1}| \cdot |DIS_{ag2}(y, y')|. \end{aligned} \quad (24)$$

By (24),

$$\begin{aligned} & \mu_{ag}(o_{ag}(x, y), o_{ag}(x', y'), t) \\ & = 1 - \frac{|DIS_{ag}(o_{ag}(x, y), o_{ag}(x', y'))|}{|A_{ag1}| \cdot |A_{ag2}|} \\ & \geq 1 - \frac{|DIS_{ag1}(x, x')| \cdot |A_{ag2}| + |A_{ag1}| \cdot |DIS_{ag2}(y, y')|}{|A_{ag1}| \cdot |A_{ag2}|} \\ & = 1 - \frac{|DIS_{ag1}(x, x')|}{|A_{ag1}|} + 1 - \frac{|DIS_{ag2}(y, y')|}{|A_{ag2}|} - 1. \end{aligned} \quad (25)$$

It follows that

$$\begin{aligned} & \text{if } \mu_{ag1}(x, x', r), \mu_{ag2}(y, y', s) \text{ then} \\ & \mu_{ag}(o_{ag}(x, y), o_{ag}(x', y')L(r, s)). \end{aligned} \quad (26)$$

Hence, $UncProp(r, s) = L(r, s)$, the value of the Łukasiewicz t-norm L on the pair (r, s) .

In consequence, the granule synthesizer $GSynt_{ag}$ can be defined in our example as,

$$\begin{aligned} & GSynt_{ag}(g_{ag1}(x, r), g_{ag2}(y, s)) \\ & = (g_{ag}(o_{ag}(x, y), L(r, s))). \end{aligned} \quad (27)$$

The definition of logic synthesizer $LSynt_{ag}$ follows directly from our assumptions,

$$LSynt_{ag}(\phi_1, \phi_2) = \phi_1 \wedge \phi_2. \quad (28)$$

Finally, we consider extensions of our logical operators of intentional logic.

We have for the extension $I(\mu_{ag})_{GSynt_{ag}(g_1, g_2)}^\vee(LSynt_{ag}(\phi_1, \phi_2)) :$

$$\begin{aligned} & I(\mu_{ag})_{GSynt_{ag}(g_1, g_2)}^\vee(LSynt_{ag}(\phi_1, \phi_2)) \\ & = I(\mu_{ag1})_{g_1}^\vee(\phi_1) \cdot I(\mu_{ag2})_{g_2}^\vee(\phi_2), \end{aligned} \quad (29)$$

which follows directly from (27), (28).

Let us note that

$$\begin{aligned} & I(\mu_{ag1})_{g_1}^\vee(\phi_1) \cdot I(\mu_{ag2})_{g_2}^\vee(\phi_2) \\ & = P\left(I(\mu_{ag1})_{g_1}^\vee(\phi_1), I(\mu_{ag2})_{g_2}^\vee(\phi_2)\right), \end{aligned}$$

where P is the Product Archimedean t-norm.

Thus, in the case of parallel fusion, where each agent works according to the Łukasiewicz t-norm, uncertainty propagation and granule synthesis are described by the Łukasiewicz t-norm L and extensions of logical intensions propagate according to the Product t-norm P .

Reasoning in Cognitive Schemes: Rough Neural Reasoning

In neural models of computation, see Bishop (1997), an essential feature of neurons is differentiability of transfer functions; hence, we introduce a special type of rough inclusions, called *Gaussian* in Polkowski (2005a) because of their form, by letting,

$$\mu_G(x, y, r) \text{ iff } e^{-\left|\sum_{a \in DIS(x, y)} w_a\right|^2} \geq r, \quad (30)$$

where $w_a \in (0, +\infty)$ is a weight associated with the attribute a for each attribute $a \in A$; clearly, we retain notation of previous sections. One may notice that the Gaussian rough inclusion is a modification of the rough inclusion μ_P obtained from the Product t-norm.

Let us observe in passing that μ_G can be factored through the indiscernibility relation $IND(A)$, and thus, its arguments can be objects as well as indiscernibility classes; we will freely use this fact.

Properties of Gaussian rough inclusions are following, cf., Polkowski (2005b):

- $x \text{ ing } y$ iff $DIS(x, y) = \emptyset$.
- There exists a function $\eta(r, s)$ such that $\mu_G(x, y, r), \mu_G(y, z, s)$ imply $\mu_G(x, z, \eta(r, s))$.
- if $x \text{ ing } g_r^{\mu_G} y, x \text{ ing } g_s^{\mu_G} z$, then $g_t^{\mu_G} x \text{ ing } g_r^{\mu_G} y, g_t^{\mu_G} x \text{ ing } g_s^{\mu_G} z$ fort $\geq \max \{r^4, s^4\}$.

Property 1 follows by definition, and Property 2 may be verified with $\eta(r, s) = r \cdot s \cdot e^{2 \cdot (\log r \cdot \log s)^{1/2}}$ (op.cit.). Property 3 can be verified by observing that t should satisfy conditions $\eta(r, t) \geq r, \eta(s, t) \geq s$ because of Property 2, see Polkowski (2005b).

Rough Mereological Perceptron The rough mereological perceptron is modeled on the perceptron, see, e.g., Bishop (1997) and it consists of an intelligent agent ag , endowed with a Gaussian rough inclusion μ_{ag} on the information system $I_{ag} = (U_{ag}, A_{ag})$ of the agent ag .

The input to ag is in the form of a finite tuple $\bar{x} = (x_1, \dots, x_k)$ of objects, and the input $\bar{x}$ is converted at ag into an object $x = O_{ag}(\bar{x}) \in U_{ag}$ by means of an operator O_{ag} .

The rough mereological perceptron is endowed with a set of *target concepts* $T_{ag} \subseteq U_{ag}/IND(A_{ag})$, each target concept a class of the indiscernibility IND_{ag} .

Formally, a rough mereological perceptron is thus a tuple

$$RMP = (ag, I_{ag}, \mu_{ag}, O_{ag}, T_{ag}).$$

The output $res_{ag}(\bar{x})$ to RMP , at the input $\bar{x}$, is a granule of knowledge $g_{r(res)}x$ with

$$r(res) = \max \{r : \text{there is } y \in T_{ag} \mu_{ag}(x, y, r)\}. \quad (31)$$

Formula (31) tells that the input $\bar{x}$ is classified at ag as the collection of in-discernibility classes that are as close to x as the closest target class (closeness meant with respect to μ_{ag}).

Applications in Data Mining: Granular Classifiers

Granulation of knowledge by means of μ_h consists in forming for each $r \in [0, 1]$ and each $u \in U$, of a granule $g_h(u, r) = \{v \in U : \mu_h(v, u, r)\}$. As, clearly, μ_h is symmetric, from $v \in g_h(u, r)$ it follows that $u \in g_h(v, r)$.

Granular data sets were proposed in Polkowski (2005a, 2006, 2007a) as the following constructions. Given $r \in [0, 1]$, the set of all granules $G_r^h = \{g_h(u, r) : u \in U\}$ is defined. From this set, a covering $Cov_r^h(\mathcal{G})$ is chosen according to a strategy $\mathcal{G}$. Granules in $Cov_r^h(\mathcal{G})$ form a new universe of objects. For each $g \in Cov_r^h(\mathcal{G})$, and each attribute $a \in A$, a factored attribute a_h is defined as $a_h(g) = S(\{a(u) : u \in g\})$.

The new information system $I_r^h = (Cov_r^h(\mathcal{G}), \{a_h : a \in A\})$ is a *granular reflection* of the original information system I . The same procedure is applied to a decision system $D = (U, A, d)$ to form the reflection $D_r^h = (Cov_r^h(\mathcal{G}), \{a_h : a \in A\}, d_h)$. The object $o(g)$ defined for a granule g by means of $\inf(o(g)) = \{(a_h, a_h(g)) : a \in A\}$ according to a strategy S is called an *S-reflection of the granule g*; clearly, $o(g)$ need not be a real object in the training or test sets.

The idea of data classification by means of granular reflections consists in splitting a given data set into the training and test parts, forming a granular reflection of training set, inducing classification rules from this new data set and applying the induced rules in classifying data in the original test set.

We include some examples showing the high efficiency of this approach, see Artiemjew (2007), Polkowski (2007a, 2011), Polkowski and Artiemjew (2007, 2015).

In tests with μ_h , the exhaustive classifier (for its public version, see Skowron et al. (1994)) was applied, cf., Artiemjew (2007), Polkowski (2007a). The strategy $\mathcal{G}$ was a random choice of an irreducible covering from the set of all granules of a given radius and the strategy $\mathcal{S}$ was selected as majority voting with random tie resolution.

We show results of tests with Australian credit data set, see UCI Repository (<http://www.ics.uci.edu/mllearn/databases/>), well studied in rough set literature: we include for comparison, some best results obtained by means of rough set-based methods, in Table 5. Classification quality is expressed by means of two factors: accuracy and coverage where which is the ratio of the number of correctly classified objects to the number of recognized test objects (called also *total accuracy*) and *total coverage*, $\frac{rec}{test}$, where *rec* is the number of recognized test cases and *test* is the number of test cases.

Tests on this data with granular approach indicated above were carried by splitting the Australian credit data set into the training and test sets in ratio 1:1; the training sample was granulated and a granular reflection was formed

from which by means of RSES exhaustive algorithm a classifier was produced which was applied to the test part of data to find quality of classification.

Granules were calculated in a twofold way: first as indicated above and second, by a modified procedure of *concept dependent granulation*, see Artiemjew (2007): in the latter procedure, the granule $g_h^c(u, r) = g_h(u, r) \cap [u]_d$ was computed relative to the *concept*, that is, decision class, to which u belonged. The results of tests are given in Table 6 in which the best results obtained with various granulation radii are shown, taken from Polkowski and Artiemjew (2007).

Results in Table 6 do witness that granular approach gives results fully comparable with other results for satisfactorily large radii of granulation, whereas the concept-dependent granulation gives results better than any other existing approach.

In order to test the impact which the choice of granular covering has on classification, we have carried out 10 experiments with the RSES exhaustive algorithm, with random coverings on the Heart dataset (Cleveland), see UCI Repository (<http://www.ics.uci.edu/mllearn/databases/>). Results are given in Table 7. Total accuracy was found to be 0.807, and total coverage 1.0 with exhaustive algorithm on the full data. These

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 5 Best results for Australian credit by some rough set based algorithms

Source	Method	Accuracy	Coverage
Bazan (1998)	<i>SNAPM</i> (0.9)	<i>error</i> = 0.130	–
Nguyen (Lin 2005)	<i>simple.templates</i>	0.929	0.623
Nguyen (Lin 2005)	<i>general.templates</i>	0.886	0.905
Nguyen (Lin 2005)	<i>tolerance.gen.templ.</i>	0.875	1.0
Wroblewski (Stefanowski 1998)	<i>adaptive.classifier</i>	0.863	–

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 6 Best results for Australian credit by granular approach

Source	Method	Accuracy	Coverage
Polkowski and Artiemjew (2015)	<i>granular radius</i> 0.642857	0.867	1.0
Polkowski and Artiemjew (2015)	<i>granular radius</i> 0.714826	0.875	1.0
Polkowski and Artiemjew (2015)	<i>granular.concept dependent radius</i> 0.785	0.9970	0.9995

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 7 Effect of a choice of a granular covering on classification

Radius	Total accuracy	Total coverage
0.0	0.0	0.0
0.0769231	0.0	0.0
0.153846	0.0	0.0
0.230769	0.0–0.789	0.0–1.0
0.307692	0.01.0	0.0–1.0
0.384615	0.737–0.799	0.993–1.00
0.461538	0.778–0.822	0.996–1.0
0.538462	0.881–0.911	1.0
0.615385	0.874–0.907	1.0
0.692308	0.963–0.974	1.0
0.769231	1.0	1.0
0.846154	1.0	1.0
0.923077	1.0	1.0
1.0	1.0	1.0

values are achieved here with radius of at least 0.538462, and beginning with the radius of 0.384615, the error in total accuracy is at most 0.07, and the error in total coverage is at most 0.007.

This does witness a very high stability of the granular approach showing the essential independence of results of a choice of a granular covering for inducing a granular reflection of data.

Granules from Parameterized Variants of Rough Inclusions μ_h in Classification of Data

Given an $\varepsilon \in [0, 1]$, we let $\mu_h^\varepsilon(v, u, r)$ if and only if $|\{a \in A : \rho(a(v), a(u)) < \varepsilon\}| \geq r \cdot |A|$. It is manifest that μ^ε is a rough inclusion if ρ is a nonarchimedean metric, that is, $\rho(u, w) \leq \max\{\rho(u, v), \rho(v, w)\}$; otherwise the monotonicity condition (RM2) of section “[Introduction: Basic Ideas](#)” need not be satisfied and this takes place with most popular metrics like Euclidean, Manhattan, etc.

In this case, a rough inclusion is μ^* defined as follows: $\mu_h^*(u, v, r)$ if and only if there exists an ε such that $\mu_h^\varepsilon(v, u, r)$. Then it is easy to check that μ^* is a rough inclusion. The parameter r is called the catch radius.

Granules induced by the rough inclusion μ_h^* with $r = 1$ have a simple structure: a granule $g_h^\varepsilon(u, 1)$ consists of all $v \in U$ such that $\rho(a(u), a(v)) \leq \varepsilon$.

The idea poses itself to use granules defined in this way to assign a decision class to an object u in the test set. First on the training set, rules are induced by an exhaustive algorithm. Then, given a set Rul of these rules, and an object u in the test set, a granule $g_h^\varepsilon(u, 1)$ is formed in the set Rul : in this, the duality between objects and rules is exploited as rules and objects can be written down in a same format of information sets. This also allows for using training objects instead of rules in forming granules and voting for decision by majority voting.

Thus, $g_h^\varepsilon(u, 1) = \{r \in Rul : \rho(a(u), a(r)) \leq \varepsilon \text{ for each attribute } a \in A(r)\}$ where $a(r)$ is the value of the attribute a in the premise of the rule and $A(r)$ is the set of attributes in the premise of r .

Rules in the granule $g_h^\varepsilon(u, 1)$ are taking part in a voting process: for each value c of a decision class, the following factor is computed,

$$\text{param}(c) = \frac{\text{sum of supports of rules pointing to } c}{\text{cardinality of } c \text{ in the training set}}. \quad (32)$$

cf., Bazan (1998), Michalski et al. (1986) for a discussion of various strategies of voting for decision values.

The class c_u assigned to u is decided by

$$param(c_u) = \max_c param(c); \quad (33)$$

with random resolution of ties.

In computing granules, the parameter ϵ is normalized to the interval $[0,1]$ as follows: first, for each attribute $a \in A$, the value $train(a) = \max_{training\ set} a - \min_{training\ set} a$ is computed and the real line $(-\infty, +\infty)$ is contracted to the interval $[\min_{training\ set} a, \max_{training\ set} a]$ by the mapping f_a ,

$$f_a(x) = \begin{cases} \min_{training\ set} a & \text{in case } x \leq \min_{training\ set} a \\ x & \text{in case } x \in [\min_{training\ set} a, \max_{training\ set} a] \\ \max_{training\ set} a & \text{in case } x \geq \max_{training\ set} a. \end{cases} \quad (34)$$

When the value $a(u)$ for a test object u is off the range $[\min_{training\ set} a, \max_{training\ set} a]$, it is replaced with the value $f_a(a(u))$ in the range. For an object v , or a rule r with the value $a(v)$, resp., $a(r)$ of a denoted $a(v, r)$, the parameter ϵ is computed as $\frac{|a(v, r) - f_a(a(u))|}{train(a)}$. The metric ρ was chosen as the metric $|x - y|$ in the real line. We show results of experiments with rough inclusions discussed in this work. Our data set was a subset of Australian credit data in which training set had 100 objects from class 1 and 150 objects from class 0 (which approximately yields the

distribution of classes in the whole data set). The test set had 100 objects, 50 from each class. The RSES exhaustive classifier (see Skowron et al. 1994) applied to this data set gave accuracy of 0.79 and coverage of 1.0.

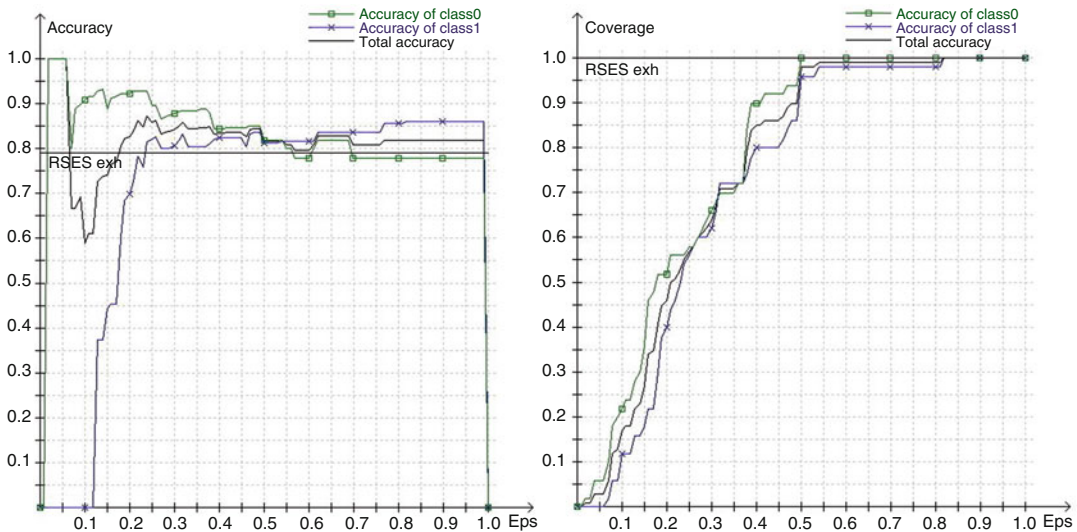
In Fig. 1 results of classification are given in function of ϵ for accuracy as well as for coverage.

We return to the rough inclusion $\mu_h^*(v, u, r)$ with general radius r . The procedure applied in case of $\mu_h^e(v, u, 1)$ can be repeated in the general setting. The resulting classifier is a function of two parameters ϵ, r . In Table 8 results are included where against values of the catch radius r (defined on p. 32) the best value for ϵ 's marked by the optimal value *optimal eps* is given for accuracy and coverage.

As shown in section “Granulation of Knowledge: General Ideas and Technique of Rough Inclusions,” residual implications of continuous t-norms can supply rough inclusions according to a general formula, see (9),

$$\mu_\phi(v, u, r) \text{ iff } \phi(u) \Rightarrow_r \phi(v) \geq r, \quad (35)$$

where ϕ maps the set U of objects into $[0, 1]$ and $\phi(u) \leq \phi(v)$ if and only if $u \text{ ing } v$ (ing is an ingredient relation of the underlying mereology;



Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Fig. 1 Results for algorithm 1_v1, Best result for $\epsilon = 0.62$: accuracy = 0.828283, coverage = 0.99

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 8 (40–60%)(1-0);

Australian credit;
Algorithm 1_v2.
r_catch=catch radius,
optimal_eps=Best ε ,
acc=accuracy, cov=
coverage

r_catch	Optimal eps	acc	cov
nil	nil	0.79	1.0
0.071428	0	0.06	1.0
0.142857	0	0.66	1.0
0.214286	0.01	0.74	1.0
0.285714	0.02	0.83	1.0
0.357143	0.07	0.82	1.0
0.428571	0.05	0.82	1.0
0.500000	0	0.82	1.0
0.571429	0.08	0.84	1.0
0.642857	0.09	0.84	1.0
0.714286	0.16	0.85	1.0
0.785714	0.22	0.86	1.0
0.857143	0.39	0.84	1.0
0.928571	0.41	0.828283	0.99
1.000000	0.62	0.828283	0.99

$\Rightarrow_t$ is the residual implication induced by the t-norm.

A weak interesting variant of this class of rough inclusions is indicated. This variant uses sets

$$dis_\varepsilon(u, v) = \frac{|\{a \in A : \rho(a(u), a(v)) \geq \varepsilon\}|}{|A|}, \text{ and}$$

$$ind_\varepsilon(u, v) = \frac{|\{a \in A : \rho(a(u), a(v)) < \varepsilon\}|}{|A|},$$

for $u, v \in U$, $\varepsilon \in [0, 1]$, where ρ is a metric $|x - y|$ on attribute value sets.

The resulting weak variant of the rough inclusion μ_ϕ is

$$\mu_t(u, v, r) \text{ iff } dis_\varepsilon(u, v) \rightarrow_t ind_\varepsilon(u, v) \geq r. \quad (36)$$

Basic variants for three principal t-norms: the Łukasiewicz t-norm $L = \max\{0, x + y - 1\}$, the product t-norm $P(x, y) = x \cdot y$, and $\min\{x, y\}$ are, (the value in all variants is 1 if and only if $x \leq y$ so we give values only in the contrary case)

$$\mu_t(u, v, r) \text{ iff } \begin{cases} 1 - dis_\varepsilon(u, v) + ind_\varepsilon(u, v) \geq r \text{ for } L \\ \frac{ind_\varepsilon(u, v)}{dis_\varepsilon(u, v)} \geq r \text{ for } P \\ ind_\varepsilon(u, v) \geq r \text{ for } \min \end{cases}$$

(37)

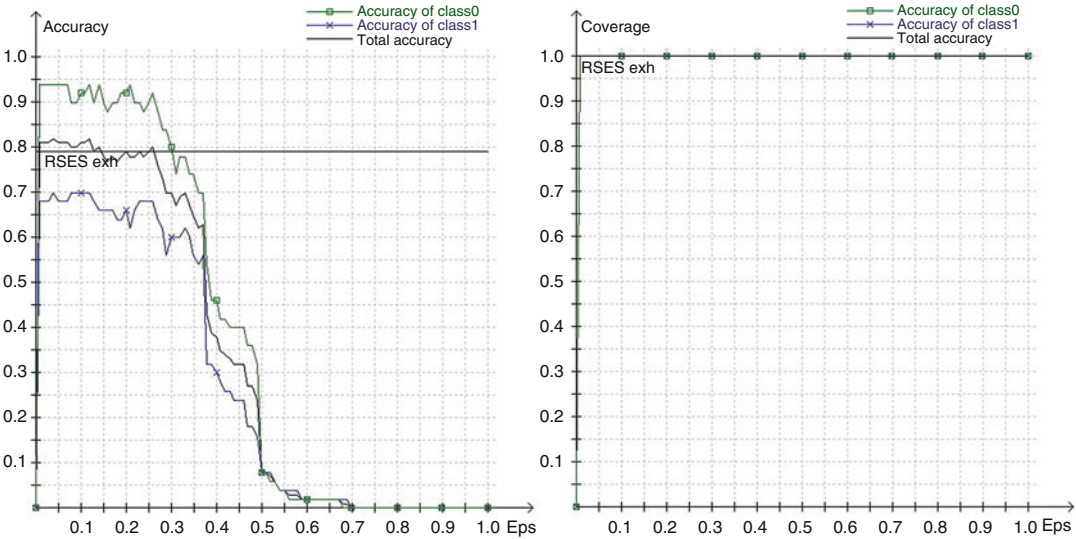
Objects in the class c in the training set vote for decision at the test object u according to the formula: $p(c) = \frac{\sum_{v \in c} w(v, t)}{|c| \text{ the training set}}$ where weight $w(v, t)$ is $dis_\varepsilon(u, v) \rightarrow_t ind_\varepsilon(u, v)$; rules induced from the training set pointing to the class c vote according to the formula $p(c) = \frac{\sum_{r \text{ in the training set}} w(r, t) \cdot support(t)}{|c|}$. In either case, the class c^* with $p(c^*) = \max p(c)$ is chosen. We include here results of tests with training objects and $t = \min$ (Fig. 2) and rules and $t = \min$ (Fig. 3).

Similarly, we include in Figs. 4 and 5 results of tests with granules of training objects and rules for $t = P$, the product t-norm.

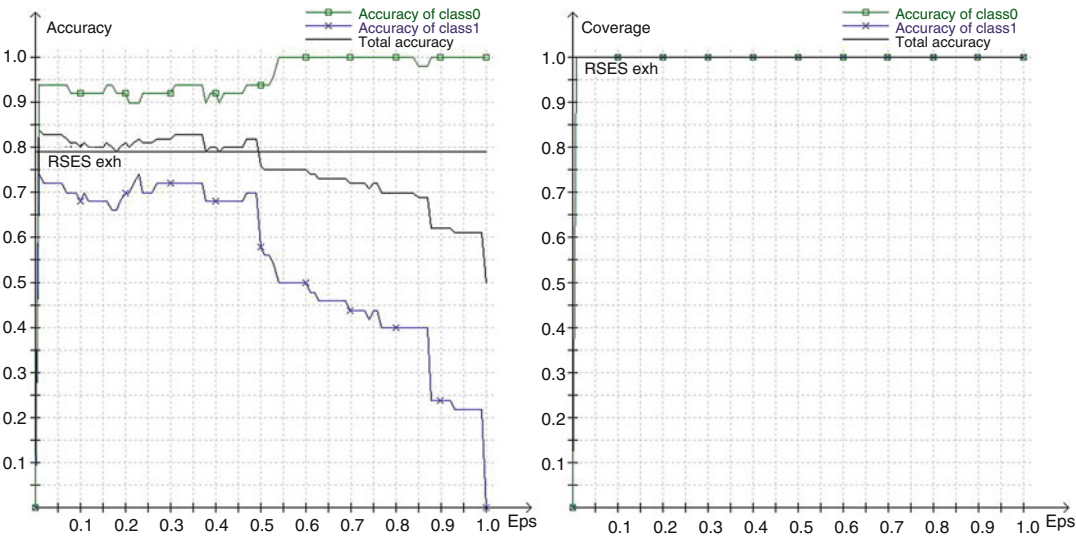
The results of tests in best cases for optimal values of ε exceed results obtained with the standard exhaustive algorithm (the RSES exhaustive algorithm gives in this data set accuracy of 0.79 and coverage 1.0).

In Polkowski and Artiemjew (2015), an extensive research has been presented on various strategies for granular computing in data mining. Authors examine there the following classification strategies

We apply the following classification algorithms.



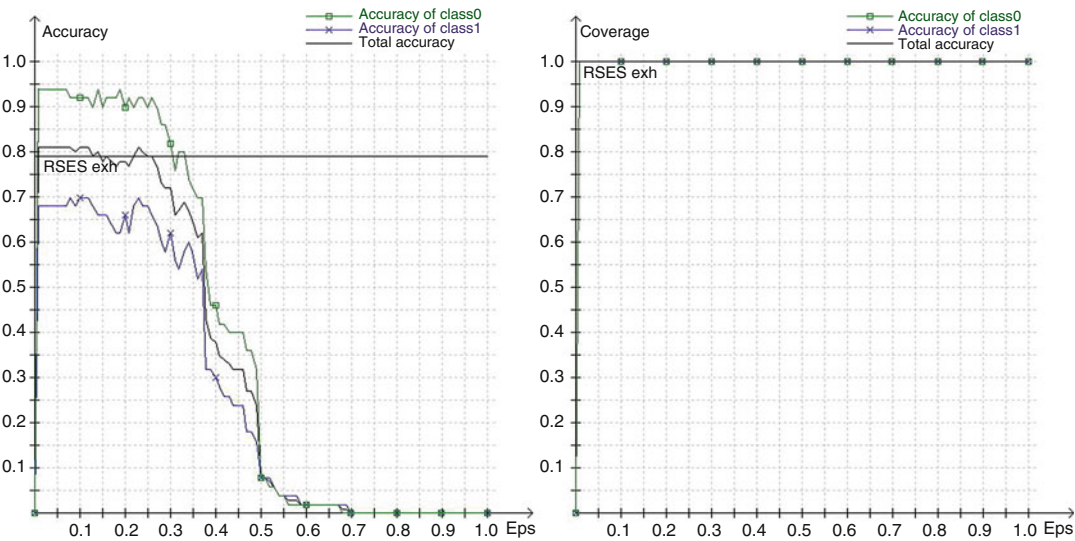
Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Fig. 2 Results for algorithm 5_v1, Best result for $\varepsilon = 0.04$, accuracy = 0.82, coverage = 1



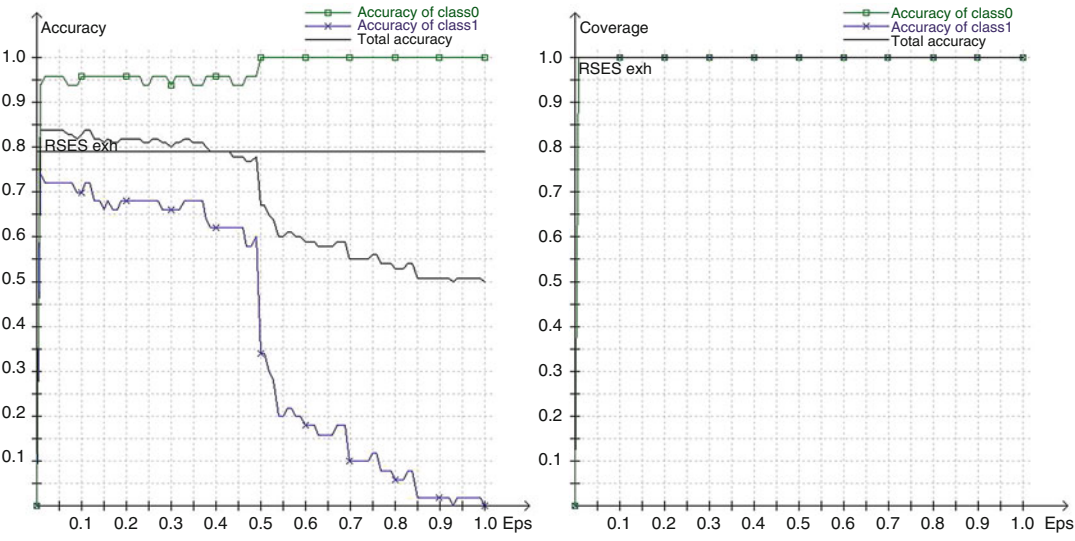
Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Fig. 3 Results for algorithm 5_v2, Best result for $\varepsilon = 0.01$, accuracy = 0.84, coverage = 1

Algorithm_1_v1: Classification by the training set
Algorithm_1_v2: Classification by the training set with the catch radius
Algorithm_2_v1: Classification by granules of the training set
Algorithm_2_v2: Classification by granules of the training set with the catch radius

Algorithm_3_v1: Classification by decision rules induced from the training set by the *exhaustive method*
Algorithm_3_v2: Classification by the decision rules *exhaustive* with the catch radius
Algorithm_4_v1: Classification by the decision rules *exhaustive* from the granular reflection



Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Fig. 4 Results for algorithm 6_v1, Best result for $\varepsilon = 0.01$, accuracy = 0.81, coverage = 1



Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Fig. 5 Results for algorithm 6_v2, Best result for $varepsilon = 0.01$, accuracy = 0.84, coverage = 1

Algorithm_4_v2: Classification by the decision rules *exhaustive* from the granular reflection with the catch radius

Algorithm_5_v1: Classification by things in the decision system weighted by the minimum residual implication

Algorithm_5_v2: Classification by the decision rules *exhaustive* from the training set weighted by the minimum residual implication

Algorithm_5_v3: Classification by granules of the training set weighted by the minimum residual implication

Algorithm_5_v4: Classification by the decision rules *exhaustive* from the granular reflection weighted by the minimum residual implication

Algorithm_6_v1: Classification by things in the decision system weighted by the product residual implication

Algorithm_6_v2: Classification by the decision rules *exhaustive* from the training set weighted by the product residual implication

Algorithm_6_v3: Classification by granules of the training set weighted by the product residual implication

Algorithm_6_v4: Classification by the decision rules *exhaustive* from the granular reflection weighted by the product residual implication

Algorithm_7_v1: Classification by things in the decision system weighted by the Łukasiewicz residual implication

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 9 Australian Credit; in ¹ the best of all result; in ²

reduction of training things is 17.6%, in ³, ⁴, ⁶ i ⁷ it is 3.4%, in ⁵ it is 3.3%

No.	Author	Method	Accuracy	Coverage
1.	Bazan[?]	SNAPM (0.9)	0.870	–
2.	Statlog[?]	Standard (NNANR)	0.860	–
3.	Statlog[?]	The best dynamic method	813	–
4.	Statlog[?]	The best RS method	0.870	–
5.	literature	Cal5 CV – 10	0.869	–
6.		k – NN	0.819	–
7.		C4.5	0.845	–
8.	S.H.Nguyen[?]	simple.templates	0.929	0.623
9.	S.H.Nguyen[?]	general.templates	0.886	0.905
10.	S.H.Nguyen[?]	closest.simple.templates	0.821	1.0
11.	S.H.Nguyen[?]	closest.gen.templates	0.855	1.0
12.	S.H.Nguyen[?]	tolerance.simple.templ.	0.842	1.0
13.	S.H.Nguyen[?]	tolerance.gen.templ.	0.875	1.0
14.	J.Wroblewski[?]	adaptive.classifier	0.863	–
	5 – fold Cross–Validation			
14.	our result	Alg_8_v1_wariant5 ¹ . $\epsilon_{opt} = 0.13$	0.880	1.0
15.	our result	Alg_2_v2 ² . $r_{gran} = 0.714286$. $r_{catch} = 0.714286$. $\epsilon_{opt} = 0.08$	0.875	1.0
16.	our result	Alg–1.v2. $r_{catch} = 0.785714$. $\epsilon_{opt} = 0.18$	0.872	1.0
17.	our result	Alg–3–v1. $\epsilon_{opt} = 0.46$	0.871	1.0
18.	our result	Alg–3.v2. $r_{catch} = 0.142857$. $\epsilon_{opt} = 0.35$	0.868	1.0
19.	our result	Alg_2_v1 ³ . $r_{gran} = 0.785714$. $\epsilon_{opt} = 0.54$	0.862	0.996
20.	our result	Alg_5_v2. $\epsilon_{opt} = 0.02$	0.861	1.0
21.	our result	Alg–1–v1. $\epsilon_{opt} = 0.83$	0.859	0.999
22.	our result	Alg_7_v3 ⁴ . $r_{gran} = 0.785714$. $\epsilon_{opt} = 0.01$	0.859	1.0
23.	our result	Alg_5_v3 ⁵ . $r_{gran} = 0.785714$. $\epsilon_{opt} = 0.05$	0.855	1.0
24.	our result	Alg_6–v3 ⁶ . $r_{gran} = 0.785714$. $\epsilon_{opt} = 0.01$	0.852	1.0
25.	our result	Alg_4.v1 ⁷ . $r_{gran} = 0.785714$, $\epsilon_{opt} = 0.01$	0.851	1.0
26.	our result	Alg_6_v2. $\epsilon_{opt} = 0.01$	0.851	1.0
27.	our result	Alg_5_v1. $\epsilon_{opt} = 0.04$	0.848	1.0
28.	our result	Alg_6_v1. $\epsilon_{opt} = 0.06$	0.848	1.0
29.	our result	Alg_7_v1. $\epsilon_{opt} = 0.05$	0.846	1.0
30.	our result	Alg_7_v2. $\epsilon_{opt} = 0$	0.555	1.0

Algorithm_7_v2: Classification by the decision rules *exhaustive* from the training set weighted by the Łukasiewicz residual implication

Algorithm_7_v3: Classification by granules of the training set weighted by the Łukasiewicz residual implication

Algorithm_7_v4: Classification by the decision rules *exhaustive* from the granular reflection weighted by the Łukasiewicz residual implication

Algorithm_8_v1_variants 1,2,3,4,5: Classification is by weighted voting of things in the training set

The variant to the above called *the ε granulation* rests in replacing the condition $a(u) = a(v)$ with the condition $|a(u) - a(v)| \leq \varepsilon$.

In Table 9 we match results by rough set methods obtained by other authors with results obtained by us for algorithms proposed above.

The best result by algorithms of series 1 to 8 is obtained by Algorithm 8_1** (line 14) for the optimal value of $\varepsilon = .13$ with accuracy .880 and coverage of 1.0. Slightly worse accuracy was obtained by Algorithms 2_v2, 1_v2, 3_v1, 3_v2 (lines 15–18) still with coverage = 1.0.

The best accuracy by other methods was obtained by means of the simple (.929) as well as general (.886) template approach (lines 8, 9) but with much smaller coverage of, respectively, .629 (less than two thirds of data covered) and 0.905.

In Table 10, we collect the results for Australian credit obtained by other methods.

Granulation of Knowledge: Similarity Based Approach in Information and Decision Systems, Table 10 Australian Credit; best classification results from Statlog [?]

Source	Method	Accuracy	Coverage
Statlog[?]	10 – fold Cross–Validation Cal5	0.869	–
KG	K – NN, k = 18, manh, std	0.864	–
Statlog[?]	ITrule	0.863	–
KG	w – NN, k = 18, manh, simplex, std	0.862	–
Statlog[?]	Discrim	0.859	–
Statlog[?]	DIPOL92	0.859	–
Statlog[?]	C4.5	0.855+ / –0.007	–
Statlog[?]	CART	0.855	–
Statlog[?]	RBF	0.855	–
Statlog[?]	CASTLE	0.852	–
Statlog[?]	NaiveBay	0.849	–
Statlog[?]	IndCART	0.848	–
KG	k – NN, k = 11, std, eucl	0.848	–
Statlog[?]	Backprop	0.846	–
Statlog[?]	C4.5	0.845	–
KG	k – NN, k = 11, fec.sel, eucl, std	0.844	–
Statlog[?]	SMART	0.842	–
Statlog[?]	Baytree	0.829	–
Statlog[?]	k – NN	0.819	–
Statlog[?]	NewID	0.819	–
Statlog[?]	AC2	0.819	–
Statlog[?]	LVQ	0.803	–
Statlog[?]	ALLOC80	0.799	–
Statlog[?]	CN2	0.796	–
Statlog[?]	Quadisc	0.793	–
Statlog[?]	Default	0.760	–

Complexity Issues

Concerning granulation in information systems, it can be proved that finding a granule about an object u is at most $O(|U| \cdot |A|)$. Therefore, finding all granules of the given radius is $O(|U|^2 \cdot |A|)$.

Selecting a coverage by granules of the given radius is therefore $O(|U|^2 \cdot |A|)$.

Future Directions

As results of the last section witness, granulation is a very important ingredient in such diverse areas as reasoning about uncertainty, reasoning in distributed and cognitive schemes, and data mining. Especially in this last area, it is possible to judge the contribution given by granulation objectively, by comparing quality of classifiers using granulation with those which dispense with granulation. These results show clearly that granulation produces results as a rule better than standard approach of rule induction by exhaustive algorithm and the best results are as good as any obtained by methods based on rule building by descriptors also with optimization.

Further research should be aimed at better understanding of the role of granulation, of the trade-off between the size of granules and quality of reasoning/classification, of the dynamics of granule formation and at finding better similarity measures adapted to particular data sets.

Bibliography

Primary Literature

- Artiemjew P (2007) Classifiers from granulated data sets: concept dependent and layered granulation. In: Proceedings RSKD'07. Workshop at ECML/PKDD'07. Warsaw University Press, Warsaw, pp 1–9
- Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds) (2004) The description logic handbook: theory, implementation and applications. Cambridge University Press, Cambridge, UK
- Bazan JG (1998) A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery 1. Physica Verlag, Heidelberg, pp 321–365
- Bazan JG, Son NH, Hoa NS, Synak P, Wróblewski J (2000) Rough set algorithms in classification problems. In: Polkowski L, Tsumoto S, Lin TY (eds) Rough set methods and applications. New developments in knowledge discovery in information systems. Physica Verlag, Heidelberg, pp 49–88
- Bishop CM (1997) Neural networks for pattern recognition, 2nd edn. Clarendon Press, Oxford, UK
- Greco S, Matarazzo B, Słowiński R (1999) On joint use of indiscernibility, similarity and dominance in rough approximation of decision classes. In: Proceedings of the 5th international conference of the decision sciences institute. Athens, pp 1380–1382
- Grzymala-Busse JW (1992) LERS – a system for learning from examples based on rough sets. In: Słowiński R (ed) Intelligent decision support: handbook of advances and applications of the rough sets theory. Kluwer, Dordrecht, pp 3–18
- Grzymala-Busse JW (2004) Data with missing attribute values: Generalization of indiscernibility relation and rule induction. In: Transactions on rough sets, volume 1. Lecture notes in computer science, vol 3100. Springer, Berlin, pp 78–95
- Grzymala-Busse JW, Hu M (2000) A comparison of several approaches to missing attribute values in data mining. Lecture notes in AI, vol 2005. Springer, Berlin, pp 378–385
- Klößgen W, Żytkow J (eds) (2002) Handbook of data mining and knowledge discovery. Oxford University Press, Oxford, UK
- Krawiec K et al (1998) Learning decision rules from similarity based rough approximations. In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery 2. Physica Verlag, Heidelberg, pp 37–54
- Kuratowski K, Mostowski A (1966) Set theory. Polish Scientific Publishers, Warsaw
- Leśniewski S (1916) Podstawy Ogólnej Teorii Mnogosci (On the foundations of set theory), in Polish. The Polish scientific circle, Moscow, 1916; see also a later digest: Topoi, vol 2, 1982, pp 7–52.
- Lin TY (1988) Neighborhood systems and relational database. Abstract. In: Proceedings of CSC'88, p 725
- Lin TY (1989) Neighborhood systems and approximation in database and knowledge based systems. In: Proceedings of the 4th international symposium on methodologies of intelligent systems, Poster session, pp 75–86. [download]
- Lin TY (1992) Topological and fuzzy rough sets. In: Słowiński R (ed) Intelligent decision support-handbook of applications and advances of the rough sets theory. Kluwer Academic, Dordrecht, pp 287–304
- Lin TY (1997) From rough sets and neighborhood systems to information granulation and computing with words. In: Proceedings of the European congress on intelligent techniques and soft computing, pp 1602–1606.
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh LA, Kacprzyk J (eds) Computing with words in information/intelligent systems 1. Physica Verlag, Heidelberg, pp 183–200

- Lin TY (2003) Granular computing. In: Lecture notes in computer science, vol 2639. Springer, Berlin, pp 16–24
- Lin TY (2005) Granular computing: Examples, intuitions, and modeling. In: Proceedings of IEEE 2005 conference on granular computing GrC05, Beijing, China. IEEE Press, Beijing, pp 40–44
- Lin TY (2006) A roadmap from rough set theory to granular computing. In: Proceedings RSKT 2006, 1st international conference on rough sets and knowledge technology, Chongqing, China, 2006. Lecture notes in artificial intelligence, vol 4062. Springer, Berlin, pp 33–41
- Ling C-H (1965) Representation of associative functions. *Publ Math Debr* 12:189–212
- Michalski RS et al (1986) The multi-purpose incremental learning system AQ15 and its testing to three medical domains. In: Proceedings of AAAI-86. Morgan Kaufmann, San Mateo, pp 1041–1045
- Nguyen SH (2000) Regularity analysis and its applications in Data Mining. In: Polkowski L, Tsumoto S, Lin TY (eds) *Rough set methods and applications. New developments in knowledge discovery in information systems*. Physica Verlag, Heidelberg, pp 289–378
- Nieminen J (1988) Rough tolerance equality and tolerance black boxes. *Fund Inform* 11:289–296
- Novotny M, Pawlak Z (1988) Partial dependency of attributes. *Bull Pol Acad Sci Math* 36:453–458
- Novotny M, Pawlak Z (1992) On a problem concerning dependence spaces. *Fund Inform* 16:275–287
- Pawlak Z (1982) Rough sets. *Int J Comput Inf Sci* 11:341–356
- Pawlak Z (1985) On rough dependency of attributes in information systems. *Bull Pol Acad Sci Tech* 33:551–559
- Pawlak Z (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer, Dordrecht
- Pawlak Z, Skowron A (1993) A rough set approach for decision rules generation. In: Proceedings of IJCAI'93 workshop W12. The management of uncertainty in AI, France, 1993; also ICS research report 23/93, Warsaw University of Technology, Institute of Computer Science.
- Pawlak Z, Skowron A (1994) Rough membership functions. In: Yaeger RR, Fedrizzi M, Kasprzyk J (eds) *Advances in the Dempster–Shafer theory of evidence*. Wiley, New York, pp 251–271
- Poincaré H (1905) *Science and hypothesis*. Walter Scott Publishing, London
- Polkowski L (2003) A rough set paradigm for unifying rough set theory and fuzzy set theory. In: Proceedings RSFDGrC03, Chongqing, China, 2003. Lecture notes in AI, vol 2639. Springer, Berlin, pp 70–78; also in: *Fundamenta Informaticae*, Vol. 54, pp 67–88.
- Polkowski L (2004a) A note on 3-valued rough logic accepting decision rules. *Fund Inform* 61:37–45
- Polkowski L (2004b) Toward rough set foundations. Mereological approach. In: Proceedings RSCTC04, Uppsala, Sweden. Lecture notes in AI, vol 3066. Springer, Berlin, pp 8–25
- Polkowski L (2005a) Formal granular calculi based on rough inclusions. In: Proceedings of IEEE 2005 conference on granular computing GrC05. IEEE Press, Beijing, pp 57–62
- Polkowski L (2005b) Rough-fuzzy-neurocomputing based on rough mereo-logical calculus of granules. *Int J Hybrid Intell Syst* 2:91–108
- Polkowski L (2006) A model of granular computing with applications. In: Proceedings of IEEE 2006 conference on granular computing GrC06. IEEE Press, Atlanta, pp 9–16
- Polkowski L (2007a) Granulation of knowledge in decision systems: the approach based on rough inclusions. The method and its applications. In: Proceedings RSEiSP'07 (in memory of Z. Pawlak). Lecture notes in AI, vol 4585. pp 69–79
- Polkowski L (2007b) The paradigm of granular rough computing. In: Proceedings ICCI'07. 6th IEEE international conference on cognitive informatics. IEEE Computer Society, Los Alamitos, pp 145–163
- Polkowski L, Artiemjew P (2007) On granular rough computing: factoring classifiers through granular structures. In: Proceedings RSEiSP'07, Warsaw. Lecture notes in AI, vol 4585. pp 280–289
- Polkowski L, Semeniuk-Polkowska M (2005) On rough set logics based on similarity relations. *Fund Inform* 64:379–390
- Polkowski L, Skowron A (1994) Rough mereology. In: Proceedings of ISMIS'94. Lecture notes in AI, vol 869. Springer, Berlin, pp 85–94
- Polkowski L, Skowron A (1997) Rough mereology: a new paradigm for approximate reasoning. *Int J Approx Reason* 15(4):333–365
- Polkowski L, Skowron A (1999) Towards an adaptive calculus of granules. In: Zadeh LA, Kacprzyk J (eds) *Computing with words in information/intelligent systems*, 1. Physica Verlag, Heidelberg, pp 201–228
- Polkowski L, Skowron A, Żytkow J (1994) Tolerance based rough sets. In: Lin TY, Wildberger M (eds) *Soft computing: rough sets, fuzzy logic, neural networks, uncertainty management, knowledge discovery*. Simulation Councils Inc, San Diego, pp 55–58
- Skowron A (1993) Boolean reasoning for decision rules generation. In: Komorowski J, Ras Z (eds) *Proceedings of ISMIS'93*. Lecture notes in AI, vol 689. Springer, Berlin, pp 295–305
- Skowron A, Rauszer C (1992) The discernibility matrices and functions in decision systems. In: Słowiński R (ed) *Intelligent decision support. Handbook of applications and advances of the rough sets theory*. Kluwer, Dordrecht, pp 311–362
- Skowron A, Stepaniuk J (2001) Information granules: towards foundations of granular computing. *Int J Intell Syst* 16:57–85
- Skowron A, Swiniarski RW (2004) Information granulation and pattern recognition. In: Pal SK, Polkowski L, Skowron A (eds) *Rough – neural computing. Techniques for computing with words*. Springer, Berlin, pp 599–636

- Skowron A et al (1994) RSES: a system for data analysis. <http://logic.mimuw.edu.pl/rses/>
- Stefanowski J (1998) On rough set based approaches to induction of decision rules. In: Polkowski L, Skowron A (eds) *Rough sets in knowledge discovery 1*. Physica Verlag, Heidelberg, pp 500–529
- Stefanowski J (2007) On combined classifiers, rule induction and rough sets. In: *Transactions on rough sets, volume VI. Lecture notes in computer science*, vol 4374. Springer, Berlin, pp 329–350
- UCI Repository. <http://www.ics.uci.edu/mlearn/databases/>
- van Benthem J (1988) *A manual of intensional logic*. CSLI Stanford University, Stanford
- Wille R (1982) Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival I (ed) *Ordered sets*. Reidel, Dordrecht, pp 445–470
- Wojna A (2005) Analogy-based reasoning in classifier construction. In: *Transactions on rough sets, volume IV. Lecture notes in computer science*, vol 3700. Springer, Berlin, pp 277–374
- Wróblewski J (2004) Adaptive aspects of combining approximation spaces. In: Pal SK, Polkowski L, Skowron A (eds) *Rough – neural computing. Techniques for computing with words*. Springer, Berlin, pp 139–156
- Yao YY (2000) Granular computing: basic issues and possible solutions. In: *Proceedings of the 5th joint conference on information sciences I*. Association for Intelligent Machinery, Atlantic, pp 186–189
- Yao YY (2005) Perspectives of granular computing. In: *Proceedings of IEEE 2005 conference on granular computing GrC05*. IEEE Press, Beijing, pp 85–90
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta M, Ragade R, Yeager RR (eds) *Advances in fuzzy set theory and applications*. North-Holland, Amsterdam, pp 3–18
- Zeeman EC (1965) The topology of the brain and the visual perception. In: Fort MK (ed) *Topology of 3-manifolds and selected topics*. Prentice Hall, Englewood Cliffs, pp 240–256

Books and Reviews

- Bocheński JM (1954) *Die Zeitgenössischen Denkmethode*. A. Francke AG Verlag, Bern. (Swiss Fed.)
- Duda RO, Hart PE, Stork DG (2001) *Pattern classification*. Wiley, New York
- Polkowski L (2002) *Rough sets. Mathematical foundations*. Physica Verlag, Heidelberg
- Polkowski L (2011) *Approximate reasoning by parts. An introduction to rough mereology*. ISRL vol 20. Springer, Berlin/Heidelberg. <https://doi.org/10.1007/978-3-642-22279-5>
- Polkowski L, Artiemjew P (2015) Granular computing in decision approximation. An application of rough mereology. ISRL vol 20. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-12880-1>
- Russell SJ, Norvig P (2003) *Artificial intelligence. A modern approach*, 2nd edn. Prentice Hall/Pearson Education, Upper Saddle River

Multi-Granular Computing and Quotient Structure

Ling Zhang¹ and Bo Zhang²

¹Artificial Intelligence Institute, Anhui University, Hefei, Anhui, China

²Department of Computer Science, State Key Lab of Intelligent Technology and Systems, Tsinghua University, Beijing, China

Article Outline

Glossary

Definition of the Subject

Introduction

The Granularity Relation

Hierarchy

Combination

Future Directions

Bibliography

Glossary

Granularity Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or system.

Multi-granular computing Humans are good at viewing and solving a problem at different grain-sizes (abstraction levels) and translating from one abstraction level to the others easily. This is one of basic characteristics of human intelligence. The aim of multi-granular computing is intended to investigate the granulation problem in human cognition and endow computers with the same capability to make them more efficient in problem solving.

Quotient set Given a universe X and an equivalence relation R on X , define a set $[X] = \{[x] \mid x \in X\}$, $[x] = \{y \mid y \sim x, y \in X\}$. $[X]$ is called a quotient set with respect to R , or simply a quotient set.

Quotient space Given a topologic space (X, T) , T is a topology on X and R is an equivalence relation on X . Define a quotient structure on $[X]$ as $[T] = \{u \mid p^{-1}(u) \in T, u \subset [X]\}$, where $p : X \rightarrow [X]$ is a natural projection from X to $[X]$. Construct a topologic space $([X], [T])$. Space $([X], [T])$ is a quotient space corresponding to R . There are authors who keep the neighborhood system structured but remove the axioms of topology (Lin 1989, 1992).

Quotient space model The quotient space model is a mathematical model to represent a problem at different grain-sizes by using the concept of quotient space in algebra. In the model a problem (or a system) is described by a triple (X, T, f) , with universe (domain) X , structure T and attribute f . If X represents the universe composed of the objects with the finest grain-size, when we view the same universe X at a coarser grain-size, we have a coarse-grained universe denoted by $[X]$. Then we have a new problem space $([X], [T], [f])$, where $[X]$ is the quotient universe of X , $[T]$ the corresponding quotient structure and $[f]$ the quotient attribute. The coarse space $([X], [T], [f])$ is called a quotient space of space (X, T, f) . Therefore, a problem with different grain-sizes can be represented by a family of quotient spaces.

Definition of the Subject

On one hand, any system in the world, either natural or artificial, has a multi-granular structure. This is called structural granulation. On the other hand, man always conceptualizes the world at different granularities and handles it hierarchically. This is called cognitive granulation. We believe the granulation in cognition underlies the human power in problem solving. Unfortunately, computers are generally capable of dealing with problems in only one abstraction level so far. The

motivation of the research on multi-granular computing is intended to investigate the granulation both in human cognition and real world and endow the computers with the same capacity. This will greatly reduce the computational complexity in the computerized problem solving. This strategy can be used to improve many algorithms in broad areas such as planning, search, and machine learning.

The key to the multi-granular computing is to construct a mathematical model for representing a problem at different grain-sizes. We present an algebraic model (quotient space model) of granulation that is used to analyze both human cognition and real world, especially human problem solving behaviors. The model was developed in order to represent granules and compute with them easily.

Introduction

Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or system. This term has been used in astronomy, physics, geography, photography, and information science and technology frequently. When a system is divided into components, it's important to get the right degree of componentization. The fine-grained components give much more details in constructing precisely the right functionality of a system. The coarse-grained components are easier to manage but may lose some important details. Performance and management considerations tend to favor the use of multi-granularity. For example, a text contains several levels such as chapter, paragraph, sentence, and word; each has a different grain-size. A video has scene, shot, and frame. A country may be split into state, city, and community. Man always conceptualizes an object (or system) at different granularities and deals with it hierarchically. This is one of man's characteristics that underlies his/her power. The aim of multi-granular computing is intended to investigate the granulation problem in human cognition and endow computers with the same capability. So multi-granular computing has widely been

investigated for a long time. In the data and knowledge management research community, the formalization of the concept of time granularity and its applications were addressed. A time granularity was defined by a countable set of granules; each granule can be composed of a single instant, a set of contiguous instants (time-interval), or even a set of non-contiguous instants (Bettini et al. 1998). The paper (Bettini et al. 2002) investigates the algorithm, its computational complexity and several optimization techniques to the temporal CSP (Constraint Satisfaction Problem) involving constraints in terms of different time granularities. In image processing, a multi resolution model is used extensively and it is a model which captures a wide range of levels of detail of an object and which can be used to reconstruct any one of those levels on demand. The overall goal of multiresolution modelling is to use information from objects with different spatial granularities or extract the details from complex models that are necessary for rendering a scene and to get rid of the other, unnecessary details (Iyengar and Prasad 1997; Garland 1999). In GIS applications, the representation of multi-granular spatio-temporal objects in commercially available DBSMs was addressed (Bertino et al. 2005). The research works above mostly focused on the specific application domain and lacks of a general theoretical framework. The key to the investigation is to construct a multi-granular mathematical model for a problem. Recently, there have been several models to deal with the issue such as fuzzy set (Zadeh 1997), rough set (Pawlak 1998), and others (Lin 2001; Nguyen et al. 2001; Bargiela and Pedrycz 2002). Each has its own characteristics. We presented a new model called quotient-space model (Zhang and Zhang 1992, 2004). In the model a problem (or a system) is described by a triple (X, T, f) . Universe (or domain) X represents the whole objects of the problem that we intended to deal with. It may be a point set and its power may either be finite or infinite. T is the structure of X and represents the relationship among objects of X . Structure T may have different forms but will be described by a topology on X in the following discussion. Attribute f is a set of functions defined on universe X and $f : X \rightarrow Y$ may be multi-

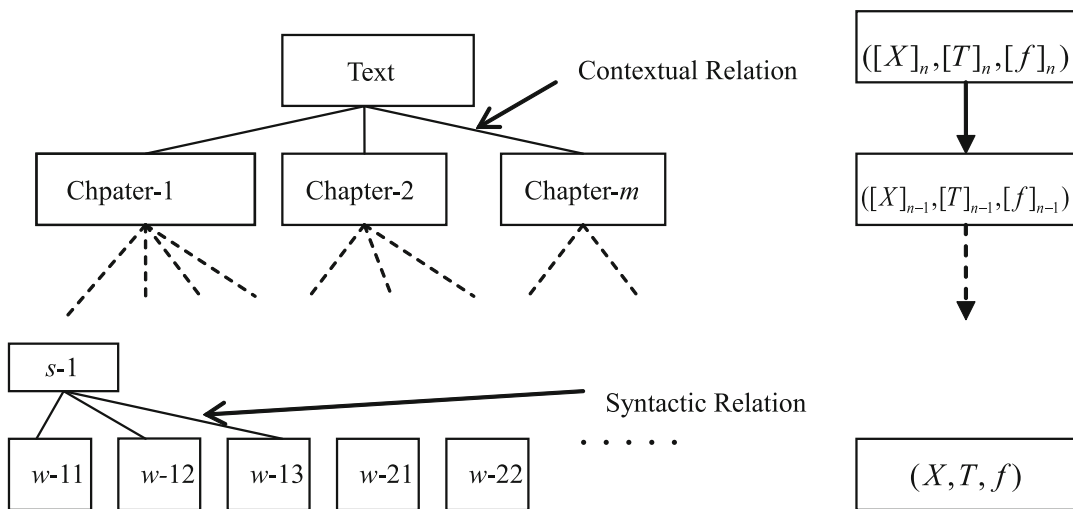
component such as $f = (f_1, f_2, \dots, f_n), f_i: X \rightarrow Y_i$, where Y_i is either a set of real numbers or other kinds of sets. Triple (X, T, f) is called a problem space (or simply a space). We will focus on the granularity relationships.

Suppose that X represents the universe composed of the components with the finest grain-size; each component is regarded as indecomposable. When we view the same universe X at a coarser grain-size, we have a coarse-grained universe denoted by $[X]$. Then we have a new problem space $([X], [T], [f])$, where $[X]$ is the quotient universe of X , $[T]$ is the corresponding quotient structure and $[f]$ the quotient attribute. For example, text X consists of a set of words. If “word” is regarded as an indecomposable component, then X is the finest universe. T is the text structure, i.e., the syntactic relationship among words. f is the property of words. If considering “sentence” as a new component, then we have a new (quotient) universe $[X]$ composed of a set of sentences; each sentence consists of a set of words. Then $[X]$ is a coarse grain-size universe. The quotient structure $[T]$ represents the contextual relationship among sentences. Quotient attribute $[f]$ represents the property of each sentence $[x]$. Then, space $([X], [T], [f])$ represents the same text with sentences as its components. The coarse space $([X], [T], [f])$ is called a quotient space of (X, T, f) (Fig. 1).

The coarse universe $[X]$ can be defined in many ways. Generally, universe $[X]$ is defined by an equivalence relation R on X . Then $[X]$ consists of the whole equivalence classes obtained by R ; each equivalence class in X is regarded as a component in $[X]$. Universe $[X]$ can also be defined by fuzzy relation and tolerance relation. In these cases, the components may have blurry boundaries or they may overlap each other.

Assume $\mathfrak{R}$ is a family of equivalence relations on X . Define an order relation ‘ $<$ ’ on $\mathfrak{R}$ as follows. Assume $R_1, R_2 \in \mathfrak{R}$; then, $R_2 < R_1$ if only if for each pair $x, y \in X$, if xR_1y , then xR_2y , where xRy indicates that x and y are R -equivalent. It implies that the universe X_1 corresponding to R_1 is finer than the universe X_2 corresponding to R_2 . So the family of quotient spaces defined by $\mathcal{R}$ is a proper mathematical model of the granulation in cognition.

A problem is represented at different granularities in human cognition. As mentioned before, its corresponding mathematical model is a family of quotient spaces; each describes the problem at a certain abstraction level. The main characteristic of our model is that the structure T is represented at the model explicitly. Based on the ‘structure’ we investigate the relation and translation among different grain-size spaces. This will facilitate the multi-granular computing and reduce the computational complexity.



Multi-Granular Computing and Quotient Structure, Fig. 1 A text represented at different grain-sizes

The Granularity Relation

The granularity relation can be represented by two basic operations on different grain-size spaces. First, the decomposition operation is the translation of a problem from a coarse level to a fine one. In the crisp-granulation issue, the universe is divided by equivalence relations, i.e., the partition. Certainly, the decomposition can be extended to fuzzy-granulation, tolerance relation, etc. In those cases, either the boundaries among granules are blurry or there is a superposition among granules. By a set of decompositions, we have a set of refined versions of the problem. Second, the projection operation is the translation from a fine level to a coarse one. By a set of projective operations; each translates a fine-grained space to a coarse one. Then we have a set of profiles of the problem, or a set of simplified versions of the problem. There are three basic kinds of projective operations, i.e., based on universe X , structure T or attribute f . By the two operations, both refined and simplified versions of the problem compose a multi-granular model.

Some relations (properties) among different grain-size spaces are shown below.

Truth Preserving Property

Assume a topological space (X, T) . By projection $p : (X, T) \rightarrow ([X], [T])$, a coarse space is constructed from (X, T) . Since the granule in $[X]$ is larger than that in X , the details of (X, T) are missing in space $([X], [T])$. If the properties of space (X, T) that we are interested in is still preserved in space $([X], [T])$, then we can also solve the same problem in the simplified space. The truth preserving property among different grain-size worlds will help us to simplify the problem solving.

Assume that R is an equivalence relation on X . From R , we have a quotient set $[X]$. A quotient topology $[T]$ induced from T can be defined as follows.

$$[T] = \{u | p^{-1}(u) \in T, u \subset [X]\} .$$

In addition, $p : X \rightarrow [X]$ is a natural projection and defined as follows.

$$\begin{aligned} p(x) &= [x] \\ p^{-1}(u) &= \{x | p(x) \in u\} . \end{aligned}$$

From topology (Eisenberg 1974), we have the following proposition.

Proposition 1: (Truth Preserving Property) If a problem $[A] \rightarrow [B]$ on $([X], [T], [f])$ has a solution path, and for $\forall [x]$, $p^{-1}([x])$ is connected on X . Then the corresponding problem $A \rightarrow B$ has a solution path on (X, T, f) consequentially.

Proof 1 Since $[A] \rightarrow [B]$ has a solution path on $([X], [T], [f])$, $[A]$ and $[B]$ fall on the same connected component C . Let $D = p^{-1}(C)$. We show that D is a connected set on X .

By reduction to absurdity, assume D is a union of two disjoint non-empty open closed sets D_1 and D_2 . $\forall a \in C$, $p^{-1}(a)$ is connected on X . $p^{-1}(a)$ only belongs to one of D_1 and D_2 . Therefore, D_i , $i = 1, 2$ consists of elements of $[X]$, i.e., there exist C_1 and C_2 such that $D_1 = p^{-1}(C_1)$, $D_2 = p^{-1}(C_2)$. Since D_i , $i = 1, 2$, are open closed sets, from the definition of natural map p , C_1 and C_2 are non-empty open closed sets on $[X]$ also. Since C_1 and C_2 are the partition of C , C is not a connected set. There is a contradiction. $\square$

In fact, in a topologic space, a problem solving (or reasoning) can be regarded as finding a connected set from the initial state (or promise) to the final state (conclusion), i.e., finding the connectivity of sets in the space. The proposition shows that if there is a solution (connected) path in the coarse-grained space $([X], [T])$, then there exists a solution path in its original space (X, T) . It shows that some problems can be solved in its coarse space rather than the original one. For example, the robotic motion planning is to find the collision-free paths in a geometrical space. If a simplified topologic space is constructed properly from the geometrical one, by the truth preserving property, we know that the problem can be transformed into that of finding a connected path in the topological space, as long as we only need to know if there exists any collision-free path. Certainly, finding connected paths in a topological

space is much easier than that in the geometrical one.

In order for the truth preserving property to remain between spaces $[X]$ and X , the condition that $\forall [x] \in [X], p^{-1}([x])$ is connected in X should be satisfied; that is, the connectivity of sets should be remained when the grain-size becomes finer. Sometimes, this is hard to come by, and we can obtain the following property.

Falsity Preserving Property

Proposition 2: (Falsity Preserving Property)

The natural projection $p : (X, T) \rightarrow ([X], [T])$ is a continuous mapping. If $A \subset X$ is a connected set on X , then $p(A)$ is a connected set on $[X]$.

It means that the connectivity of sets remains unchanged when the grain-size become coarser. This is easier to obtain than the previous one in real problem solving. In fact, in the quotient space model the human problem solving (or reasoning) can be treated as finding the connectivity of sets in the problem space. The proposition shows that if there is a solution (connected) path in the original space (X, T) , then there exists a solution path in its proper coarse-grained space $([X], [T])$. Conversely, in the coarse-grained space, if there does not exist a solution path, there is no solution in the original space. This is called the ‘falsity preserving’ property, i.e., ‘no-solution’ (region) property is preserved between quotient spaces. The property underlies the power of human hierarchical problem solving. If at the coarse level we do not find any solution in some regions, then there is no solution in the corresponding regions at the fine level. Therefore, the results obtained from the coarse level can guide the problem solving in the fine level effectively. In general, the coarse space is simpler than the fine one, so the computational complexity will be reduced by the hierarchical problem solving.

By using the structural (topological) relation among the quotient spaces, we have the two basic properties above. So the structure T plays an important role in our model. We cannot always have the proper properties as shown before. Sometime, we can only reach the properties to a certain degree such as in a probabilistic sense.

Hierarchy

There are two basic modes adopted by multi-granular computing, i.e., hierarchy and combination. The hierarchy means hierarchical problem solving strategy, i.e., problem solving process carries on from a coarse level to a fine one step by step. The combination means the integration of information observed from several coarse levels. The aim of hierarchical problem solving is to reduce the computational complexity by using the information from different grain-size worlds.

The Computational Complexity Under Deterministic Model

Given a problem space (X, T, f) , X is a finite set. $|X|$ is the number of components in X . If we solve the problem from (X, T, f) directly, i.e., find the goal from space (X, T, f) , the computational complexity is $c(|X|)$. Assume that $c(\cdot)$ only depends on the number of components of X and is independent of its structure or other attributes. The value of $c(\cdot)$ ranges over $[0, \infty)$. Assume that X_0 is a quotient space of X . c is a complexity function of X . Suppose the number of components of X_0 that might contain the goal to be g at most. First, we consider the simplest case: the “truth preserving property” comes into existence precisely, i.e., $g \equiv 1$. Now we have a set $X_1, X_2, \dots, X_t, X_{t+1} = X$ of quotient spaces, where X_i is a quotient space of X_{i+1} , $i = 1, 2, \dots, t$. We estimate the asymptotic property of computational complexity denoted by $c_t(|X|)$ under the hierarchical problem solving strategy with t levels. Let $|X| = e^n$.

Suppose that X_1 is a quotient space of X . We find the goal of X from X_1 . At a cost of complexity $c(|X|)$, we find a unique component $a_1 \in X_1$ that contains the goal. Assume that each equivalence class has the same number of components. Now the goal is within component a_1 of X_1 . If $|X_1| = e^{n_1}$, from $|X| = e^n$, we have $|a_1| = e^{n-n_1}$. The total complexity for solving problem X by the hierarchical strategy with two levels is

$$c_1(|X|) = c(|X_1|) + c(|a_1|) = c(e^{n_1}) + c(e^{n-n_1}).$$

Regarding a_1 as a set of X , if it's too large, then a_1 can further be decomposed. Assume Y_2 is a

quotient space of a_1 and $|Y_2| = e^{n_2}$. The total complexity for solving X with three hierarchical levels is

$$\begin{aligned} c_2(|X|) &= c(|X_1|) + c(|Y_2|) + c(|a_2|) \\ &= c(e^{n_1}) + c(e^{n_2}) + c(e^{n-n_1-n_2}). \end{aligned}$$

Now the goal is within the component a_2 .

By induction, the total complexity for solving X using hierarchical strategy with t levels is

$$\begin{aligned} c_t(|X|) &= c(e^{n_1}) + c(e^{n_2}) + \dots + c(e^{n_t}) \\ \text{where } n &= \sum_{i=1}^t n_i. \end{aligned} \quad (1)$$

In each abstraction level, all components are classified into b equivalence classes, $b > 0$, i.e., $\forall i, e^{n_i} = b$. Since $n = \sum_{i=1}^t n_i$, we have $e^n = b^t = e^{at}$, where $b = e^a$, $at = n$, $t = n/a$. Substituting $t = n/a$ into (1), we have

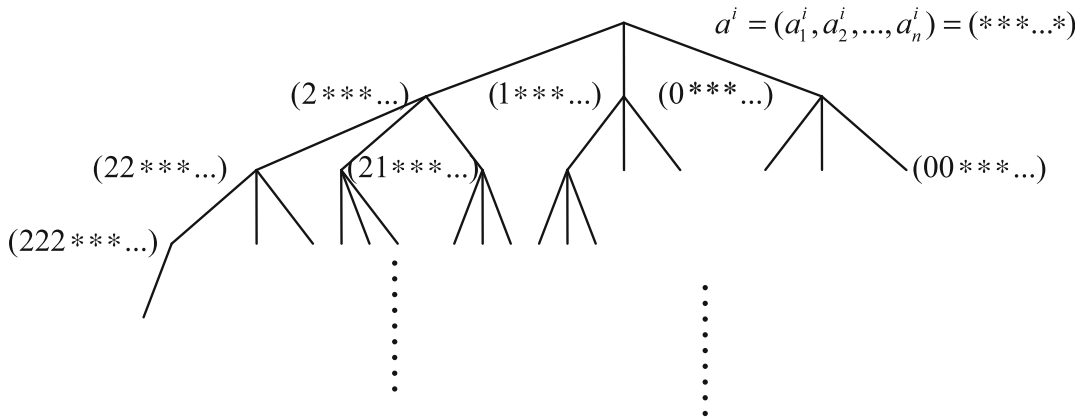
$$c_t(|X|) = t(c(b)) = c_1 n.$$

Then we have the following proposition:

Proposition 3 *If a unique component that contains the goal can be found at each abstraction level, i.e., $g \equiv 1$, there exists a hierarchical strategy such that X can be solved in a linear time ($O(n)$) in spite of the form of the original complexity $c(|X|)$.*

Example 1 We are given N balls, one of which is known to be lighter or heavier than the rest. Using a two-pan scale, how to find the counterfeit ball such that the number of weighs is minimal. First, in each weigh, we divide the balls into three groups: on the left pan, the right pan and the table denoted by '1', '2' and '0', respectively. After n weighs, each ball is assigned a ternary number with n digits denoted by $a^i = (a_1^i, a_2^i, \dots, a_n^i)$, $i = 1, 2, \dots, N$. From the quotient-space model point of view, each weigh corresponds to constructing a quotient space. After n weighs, we have a tree with depth n composed by a sequence of quotient spaces as shown in Fig. 2. Then we solve the problem hierarchically. Second, in each weigh, if the weight of balls in the left pan is lighter than that in the right pan, then denote as '1'. If the weight of balls in the right pan is lighter than that in the left one, then denote as '2'. Otherwise, i.e., when the scale balances, denote as '0'. After n weighs, we have another ternary number $e = (e_1, e_2, \dots, e_n)$ with n digits. Obviously, if the counterfeit ball is lighter, when $e_1 = 1$, the ball is annotated as '1'; when $e_1 = 2$, the ball is annotated as '2'. Similarly, when $e_1 = 0$, the counterfeit ball is annotated as '0'. Therefore, we have $a^i = e$, when the counterfeit ball is lighter.

Similarly, we have $a^i = e'$, when the counterfeit ball is heavier, where e' is a conjugate of e , i.e.,



Multi-Granular Computing and Quotient Structure, Fig. 2 The quotient-space model of ball weigh problem

$$e' = \begin{cases} 1, & e = 2 \\ 2, & e = 1 \\ 0, & e = 0 \end{cases}.$$

In order to find the counterfeit ball uniquely, the tabs cannot contain their conjugates. A ternary number with n digits has 3^n entities and $(3n + 1)/2$ pairs of conjugates. When weighing the balls for the first time, in order to make two pans have the same number of balls, the number of elements having $a_1^i = 1$ and $a_1^i = 2$ in a set $\{a^i = (a_1^i, a_2^i, \dots, a_n^i), i = 1, 2, \dots, N\}$ of tabs should be the same. So at most there are $(3^n - 1)/2$ entities of 3^n ternary numbers with n digits that can be used as tabs. Therefore, we have the following properties.

Property 1 *A counterfeit ball can be found from $(3^n - 1)/2$ balls in n weighs.*

Property 2 *A set K of tabs for weighing balls can be constructed by a set of ternary numbers with n digits that satisfies the following conditions.*

- (1) *At most one of each pair of conjugates is in set K .*
- (2) *The numbers of elements having $a_1^i = 1$ and $a_1^i = 2$ in a set $\{a^i = (a_1^i, a_2^i, \dots, a_n^i), i = 1, 2, \dots, N\}$ of tabs are the same.*

Property 3 *After n weighs, we have result e . If $e \in K, e \neq (0, 0, \dots, 0)$, the tab of the counterfeit ball is e , and the ball is lighter. If $e = (0, 0, \dots, 0)$, the tab of the counterfeit ball is e and we don't know whether it's lighter or heavier. Otherwise, if $e \notin K$, the tab of the counterfeit ball is e' and it's heavier.*

From Property 2, we know the whole optimal solution strategies of the problem. Any process with finite state can be represented by a set of binary numbers (or ternary, decimal numbers, etc.) i.e., a sequence of quotient spaces. Then the process can be managed hierarchically.

Finally, we have the following algorithm.

Algorithm 1 A set K of tabs satisfies Property 2. Each ball is assigned a tab from K . Then we have the whole tabs denoted by $a^i = (a_1^i, a_2^i, \dots, a_n^i), i = 1, 2, \dots, N$. Assume that we have result $(e_1, e_2, \dots, e_m)$ after m weighs. Taking out all balls assigned by $(e_1, e_2, \dots, e_m, a_{m+1}^i, \dots)$ and $(e'_1, e'_2, \dots, e'_m, a_{m+1}^i, \dots)$, i.e., its first m components are $(e_1, e_2, \dots, e_m)$ and $(e'_1, e'_2, \dots, e'_m)$, if $a_{m+1}^i = 1$, put the ball on the left pan; if $a_{m+1}^i = 2$, put the ball on the right pan; otherwise, $a_{m+1}^i = 0$, put it on the table. The process finishes after n weighs. If necessary, we need to put some true balls on the left and right pans such that the numbers of balls on two pans are the same.

Combination

However, in human cognition, one usually learns things from local fragments, integrates them and forms a global picture gradually. It means inferring the fine-level representation from the information collected at coarse levels. The process is called combination (or information fusion).

For a problem space (X, T, f) , given the knowledge of its two quotient spaces (X_1, T_1, f_1) and (X_2, T_2, f_2) , the 'combination' operation is intended to have an overall understanding of (X, T, f) from the known knowledge.

Let (X_3, T_3, f_3) be the combination of (X_1, T_1, f_1) and (X_2, T_2, f_2) , and $p_i : (X, T, f) \rightarrow (X_i, T_i, f_i), i = 1, 2$. In order to have a proper (X_3, T_3, f_3) , the following three combination principles should be satisfied at least.

$$\begin{cases} p_i X_3 = X_i \\ p_i T_3 = T_i \\ p_i f_3 = f_i, \quad i = 1, 2 \end{cases}$$

In general, the solution satisfying the three principles is not unique. In order to have a unique result, some criteria must be added such that the solution is optimal. First, we propose the following combination rule. Let the combination universe X_3 be the least upper bound of universes X_1 and X_2 . This implies that X_3 is the coarsest one among the universes that satisfy the first

combination principle or the finest one that we can get from the known universes X_1 and X_2 . And let the combination topology T_3 be the least upper bound of topologies T_1 and T_2 . This implies that T_3 is the coarsest one among the topologies that satisfy the second combination principle or the finest one that we can get from the known topologies T_1 and T_2 . This is also the maximal amount of information that we can obtained from the known knowledge. Therefore, the proposed X_3 and T_3 are optimal in some sense.

In most cases, the optimal criteria are domain dependent. However, here we present a general criterion as follows.

$$D(f_3, f_1, f_2) = \min_f D(f, f_1, f_2) \quad \text{or} \\ \max_f D(f, f_1, f_2).$$

Where f ranges over all attribute functions on X_3 that satisfy the third combination principle.

It's noted that in the combination principle $p_i f_3 = f_i$, $i = 1, 2$, where p_i may be a non-deterministic mapping. We'll discuss the problem in section "Combination".

To show the rationality of the above combination principles, in (Zhang and Zhang 1992) we deduced the famous Dempster–Shafer combination rule in belief theory (Shafer 1976) by the principles. It shows that the D–S rule is the outcome of the combination principles under certain optimal criteria.

Now we give an example to show how to use the combination principle.

Example 2 Semantic concept detection, which is also called high-level feature extraction in TRECVID (TREC-Video Retrieval Evaluation) (Over et al. 2005), is to automatically determine related video shots from a video dataset given some semantic concepts. This technology is very useful for automatic or semi-automatic video indexing or annotation. Usually, a small portion of data is annotated manually as training data for machine learning programs. The result of concept detection can be measured by the average precision, which summarizes the precision at all recall

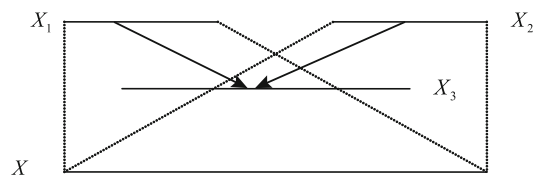
levels and favors algorithms that detect more relevant shots (Zhu 2004).

In semantic concept detection, a video shot is described by many features (attributes) with different modalities such as auto-speech recognition text, visual texture, color of segmented image regions and is related to some semantic concept. The concept detection is to determine the relevant concept of video shots from their features.

The Construction of Quotient-Space Model

A set $\{C^1, C^2, \dots, C^i, \dots, C^k\}$ of concepts and a video archive are given. Then we may extract a set of multi-granular and multi-modular features (speech, image and text) from each video shot. A quotient-space model for concept detection is constructed as follows. In the model, universe X represents a set of video shots. Based on some extracted feature of video shots, for example, the color of segmented image regions, X is classified into a set $\{a_1, a_2, \dots, a_k\}$ of classes; each corresponds to a concept C^i . Letting $X_1 = \{a_i\}$, $i = 1, 2, \dots, k$, X_1 is called a projection of X , the quotient space of X . Each class a_i consists of a set of video shots and corresponds to concept C^i . Based on another feature, for example, the auto-speech recognition text, the same X is classified into a new set $\{b_1, b_2, \dots, b_k\}$ of classes. Then, we have a new space $X_2 = \{b_j\}$, $j = 1, 2, \dots, k$ (Fig. 3).

More quotient spaces can be constructed by using different features (attributes). Then, we have a set $\{(X, T, f), (X_{n-1}, T_{n-1}, f_{n-1}), \dots, (X_0, T_0, f_0)\}$ of quotient spaces. T represents the relation among the sets of video shots. f is the description of semantic concepts of video shots that we are intended to handle. The problem is how to construct a new space X_3 based on the known information of X_1 and X_2 such that more



Multi-Granular Computing and Quotient Structure, Fig. 3 The combination of quotient spaces

relevant shots can be detected from X_3 when given a semantic concept. This is the combination problem in quotient-space model theory. Since a big gap is existed between a low-level feature and a semantic concept, the correspondence between them established by a small portion of training data has a prodigious uncertainty. To reduce the uncertainty, it generally uses the knowledge from different modalities.

The Combination of Quotient Spaces

Assume that (X_1, T_1, f_1) and (X_2, T_2, f_2) are the knowledge about $A = (X, T, f)$ in two different abstraction levels. The combination of (X_1, T_1, f_1) and (X_2, T_2, f_2) is defined as a new abstraction level of A denoted by (X_3, T_3, f_3) , which satisfies the following three basic principles.

- (1) X_1 and X_2 are quotient spaces of X_3
- (2) T_1 and T_2 are quotient structures of T_3
- (3) f_1 and f_2 are projections of f on X_1 and X_2 , respectively (X_3, T_3, f_3) might need to satisfy some other optimal criteria.

We next discuss the combination of universe X and attribute function f .

The Combination of Universes Assume that (X_1, T_1, f_1) and (X_2, T_2, f_2) are quotient spaces of (X, T, f) . R_1 and R_2 are equivalence relations with respect to X_1 and X_2 , respectively.

Define the combination universe X_3 as follows.

$xR_3 y \Leftrightarrow xR_1 y$ and $xR_2 y$, where R_3 is an equivalence relation with respect to X_3 .

It is known that all equivalence classes on X form a semi-order lattice under the relation $<$, where $R_1 < R_2 \Leftrightarrow$ if $xR_2 y$ then $xR_1 y$. In terms of the semi-order lattice, the combination of R_1 and R_2 can be defined as follows.

Definition 1 Assume that R_1 and R_2 are two equivalence relations on X . If R_3 is the least upper bound of R_1 and R_2 among the semi-order lattice, then R_3 is the combination of R_1 and R_2 . If $X_1 = \{a_i\}$ and $X_2 = \{b_j\}$ are two partitions with respect to R_1 and R_2 , respectively, then the combination of X_1 and X_2 can be represented by $X_3 = \{a_i \cap b_j \mid a_i \in X_1, b_j \in X_2\}$.

In the concept detection of video shots, X_1 is the partition of X by the speech feature and X_2 is the partition of X by the image feature, the combination universe X_3 is the partition of X by both speech feature and image feature. Obviously, X_3 is the finest universe which we can get from X_1 and X_2 . It is also the coarsest one among the universes which satisfy the above combination principle, that is, the least upper bound.

The Combination of Attribute Functions Given (X_1, T_1, f_1) and (X_2, T_2, f_2) , we find the combination space (X_3, T_3, f_3) satisfying the following conditions.

- (1) $p_i f_3 = f_i$, $i = 1, 2$, where $p_i : (X_3, T_3, f_3) \rightarrow (X_i, T_i, f_i)$ is a natural projection
- (2) $D(f, f_1, f_2)$ is a given criterion such that

$$D(f, f_1, f_2) = \min_f D(f, f_1, f_2) \text{ or } \max_f D(f, f_1, f_2).$$

Where f ranges over all attribute functions on X_3 that satisfy condition (1). In fact, the solution satisfying condition (2) is not unique. The additional optimization criteria are generally needed in order to have a unique solution. There have been several kinds of combination approaches (Maes et al. 1997; Dasarathy 2001; Jing et al. 2005), sometimes called information fusion. We show one of the possible ways.

The Combination of Attribute Functions

A set $\{x^1, x^2, \dots, x^N\}$ of video shots and a set $\{C^1, C^2, \dots, C^i, \dots, C^k\}$ of concepts are given. Using g different kinds of features such as speech, image and text, to classify the video shots by the annotated training samples, then we have g quotient spaces $\{X_1, X_2, \dots, X_g\}$. For a concept C^i , there are g different classifications $\{C_1^i, C_2^i, \dots, C_g^i\}$ on g quotient spaces, respectively; each class $C_j^i, j = 1, 2, \dots, g$, consists of a set of video shots. Assume that each sample $x^i, i = 1, 2, \dots, p$, of video shots is an identically independent distributed n -dimensional random

variable. Then class C_j^i , $j = 1, 2, \dots, g$, is a set of random variables; each can be described by a probability density function. We choose normal distribution function $N(x, \mu, \Sigma)$ as its probabilistic model, where mean μ may be defined as the center of the class, and Σ the variance matrix. Then concept C^i can be defined as the combination of C_j^i , $j = 1, 2, \dots, g$, as follows.

$$F_i(x) = \sum_j \alpha_j N(x, \mu_j, \Sigma_j), \quad j = 1, 2, \dots, g.$$

$F_i(x)$ is a describer of concept C^i . From the combination principles in quotient space model, the weights α_j can be chosen by some sort of optimization techniques. Since $F_i(x)$ is a g -component finite mixture density, the maximum likelihood estimator can be used to estimate the weights. According to the iterative EM (Expectation Maximization) algorithm presented in (Dempster et al. 1977), we have the following optimization procedures.

Let $K = \{(x^1, y^1), \dots, (x^p, y^p)\}$, $x^i \in R^n$, $y^i \in \{0, 1\}^k$ be a set of training samples. And its corresponding classification based on the set $\{C^1, C^2, \dots, C^i, \dots, C^k\}$ of concepts is denoted by $C^i = \{C_1^i, C_2^i, \dots, C_g^i\}$, $i = 1, 2, \dots, k$.

Initialization Let $\alpha_j^{(0)} = d_j$, d_j is the proportion of number of video shots in C_j^i to the total number of video shots in the i th concept.

$$\sigma_j^{(0)} = r_j, \quad r_j \text{ is the radius of } C_j^i.$$

$$\mu_j^{(0)} = a_j, \quad a_j \text{ is the center of } C_j^i.$$

$$\Sigma_j^{(0)} = (\sigma_j^{(0)})^2 I_n, \text{ where } I_n \text{ is a } n - \text{dimensional unit matrix.}$$

E Step The $(k + 1)$ -iteration, calculate the posterior probability of sample x^i from the j th component as follows.

$$\beta_{ij}^{(k)} = \beta_j(x^i, \Theta^{(k)}) = \frac{\alpha_j^{(k)} N(x^i, \mu_j^{(k)}, \Sigma_j^{(k)})}{\sum_{j=1}^g \alpha_j^{(k)} N(x^i, \mu_j^{(k)}, \Sigma_j^{(k)})},$$

$$(j = 1, \dots, g; \quad i = 1, \dots, p), \quad (2)$$

$$\alpha_j^{(k+1)} = \frac{1}{p} \sum_{i=1}^p \beta_{ij}^{(k)}, \quad j = 1, \dots, g. \quad (3)$$

M Step Find the mean and variance matrices by iteration.

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^p \beta_{ij}^{(k)} x^i}{\sum_{i=1}^p \beta_{ij}^{(k)}}, \quad \text{where } \beta_{ij}^{(k)} = \beta_j(x^i, \Theta^{(k)}),$$

$$(j = 1, \dots, g; \quad i = 1, \dots, p), \quad (4)$$

$$\Sigma_j^{(k+1)} = \frac{\sum_{i=1}^p \beta_{ij}^{(k)} (x^i - \mu_j^{(k+1)}) (x^i - \mu_j^{(k+1)})^T}{\sum_{i=1}^p \beta_{ij}^{(k)}},$$

$$(j = 1, \dots, g). \quad (5)$$

Assume $N(x, \mu, \Sigma)$ is a 1-dimensional normal distribution.

$$(\sigma_j^2)^{(k+1)} = \frac{\sum_{i=1}^p \beta_{ij}^{(k)} \left| (x^i - \mu_j^{(k+1)}) \right|^2}{\sum_{i=1}^p \beta_{ij}^{(k)}}.$$

Finally, the function $F_i(x)$ that we have is the describer (attribute) of C^i . The function $F(x) = \{F_1(x), \dots, F_p(x)\}$ is the decision function of the set $\{C^1, C^2, \dots, C^i, \dots, C^n\}$ of concepts. $F(x)$ integrates all information from g quotient spaces.

In order to show the advantage of multi-granular and multi-modal computing, Our experiments were carried out on the TRECVID 2005's test dataset for the high-level feature extraction (HFE) task (Over et al. 2005), which consists of

Multi-Granular Computing and Quotient Structure, Table 1 Comparison of uni-modal and multi-modal semantic video concept detection results

	Uni-Modal			Cross-Modal			
	ASR	Texture	Region	A + T	A + R	T + R	A + T + R
US-flag	0.0335	0.0155	0.0375	0.0359	0.0506	0.0372	0.0521
Water	0.0034	0.1143	0.0814	0.1022	0.0735	0.1333	0.1211
Mountain	0.0033	0.0693	0.1104	0.0668	0.1066	0.1176	0.1154
Sports	0.0723	0.0769	0.2156	0.1465	0.2678	0.2802	0.3050
Average	0.0281	0.0690	0.1112	0.0879	0.1246	0.1421	0.1484

86.6 h of news videos (45,766 shots in 140 video clips).

Three uni-modal methods and four multi-modal methods are compared. Uni-modal results are chosen from TRECVID 2005 submissions. They include

- *ASR*, the seventh run from Fudan University (Xue et al. 2005), based on auto-speech recognition text, which is a sub-modality of the text;
- *Texture*, the seventh run from National University of Singapore (Chua et al. 2005), based on visual texture, which is a visual sub-modality; and.
- *Region*, the first run from Tsinghua University (Yuan et al. 2005), based on color of segmented image regions, which is also a visual sub-modality.

Therefore, the four multi-modal results are denoted as ' $A + T$ ', ' $A + R$ ', ' $T + R$ ', and ' $A + T + R$ ', respectively. We use the Bayes rule of Probabilistic Model Supported Rank Aggregation (*PMSRA*) (Ding 2007) to integrate different modalities or sub-modalities. The distribution family adopted is the Gauss distribution.

Four concepts are chosen for comparison. They belong to different types of concept. '*US-flag*' is a concept of object. '*Water*' and '*Mountain*' are typical concepts of scenes. '*Sports*' is a kind of event.

Average precisions of the results based on all the uni-modal and multi-modal methods are shown in Table 1. On average, multi-modal strategies are significantly better than the uni-modal ones. Specifically, we can see that if there are no

great disparities in performances, multi-modal methods always bring significant improvement. On the other hand, if their performances are too different, which suggests the inconsistency between modalities, the average precision based on multi-modal method may be reduced a little.

Future Directions

The quotient space model that we discuss previously is defined by equivalence relations. The model should be extended to more general cases such as fuzzy relation, consistency relation. In these cases, how to construct a proper quotient structure from the original one, if the granularity relation such as "truth preserving" and "falsity preserving" properties still remains, more research works should be done in the future. Unfortunately, only a few works (Lin 2005; Zhang and Zhang 2005, 2007) dealt the problems recently.

Acknowledgments The work is supported by the National Natural Science Foundation of China under Grant. No. 60621062, and the National Key Foundation R&D Project under Grant. No.2003CB317007,2007CB311003.

Bibliography

- Bargiela A, Pedrycz W (2002) Granular computing: an introduction. Kluwer, Boston
- Bertino E, Cuadra D, Martinez P (2005) An object-relational approach to the representation of multi-granular spatio-temporal data. In: The 17th conference on advanced information systems engineering (CAiSE'05), Porto, 13–17 June, pp 119–134

- Bettini C, Wang XS, Jajodia S (2002) Solving multi-granularity temporal constraint networks. *Artif Intell* 140:107–152
- Bettini C, Dyreson CE, Evans WS, Snodgrass RT, Wang XS (1998) A glossary of time granularity concepts. In: Etzioni O, Jajodia S, Sripada S (eds) *Temporal databases: research and practice*. Lecture Notes in Computer Science, vol 1399. Springer, Berlin, pp 406–413
- Chua T-S et al (2005) TRECVID 2005 by NUS PRIS. In: TRECVID2005. NIST. <http://www-nlpir.nist.gov/projects/tvpubs/tv5.papers/nus.pdf>
- Dasarathy BV (2001) Information fusion – what, where, why, when, and how? *Ed Inform Fusion* 2(2):75–76
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data using the EM algorithm (with discussion). *J R Stat Soc SerB* 39:1–38
- Ding DY (2007) Research on information fusion technology and its application in video content analysis. Tsinghua University Dissertation (in Chinese)
- Eisenberg M (1974) *Topology*. Holt
- Garland M (1999) Multiresolution modeling: survey and future opportunities. EUROGRAGHICS'99 State of the Art Report, September 1999
- Iyengar SS, Prasad L (1997) *Wavelet analysis with applications to image processing*. CRC Press, Boca Raton
- Jing F et al (2005) A unified framework for image retrieval using keyword and visual features. *IEEE Trans Image Process* 14(7):979–989
- Lin TY (1989) Neighborhood systems and approximation in database and knowledge base systems. In: Ras ZW (ed) *Proc of the fourth international symposium on methodologies of intelligent systems*. Poster session. 12–15 Oct. North Holland, New York, pp 75–86
- Lin TY (1992) Topological and fuzzy rough sets. In: Slowinski R (ed) *Decision support by experience – application of the rough sets theory*. Kluwer, Dordrecht, pp 287–304
- Lin TY (2001) Granular fuzzy sets: a view from rough set and probability theories. *Int J Fuzzy Syst* 3(2):373–381
- Lin TY (2005) Churn-Jung Liao granular computing and rough sets. In: Maimon O, Rokach L (eds) *The data mining and knowledge discovery handbook*, Springer, pp 535–561
- Maes F et al (1997) Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16(2):187–198
- Nguyen SH, Skowron A, Stepaniuk J (2001) Granular computing: a rough set approach. *Comput Intell* 17: 514–544
- Over P, Kraaij W, Smeaton AF (2005) TRECVID 2005 – an introduction. In: TRECVID 2005, NIST. <http://www-nlpir.nist.gov/projects/tvpubs/tv5.papers/tv5intro.pdf>
- Pawlak Z (1998) Granularity of knowledge, indiscernibility, and rough sets. In: *Proc of IEEE world congress on computational intelligence*, vol 1. IEEE Press, New York, pp 106–110
- Shafer G (1976) *A mathematical theory of evidence*. Princeton University Press, Princeton
- Xue X et al (2005) Fudan University at TRECVID 2005. In: TRECVID2005. NIST. <http://www-nlpir.nist.gov/projects/tvpubs/tv5.papers/Fudan.pdf>
- Yuan J et al (2005) Tsinghua University at TRECVID 2005. In: TRECVID2005. NIST. <http://www-nlpir.nist.gov/projects/tvpubs/tv5.papers/tsinghua.pdf>
- Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 19:111–127
- Zhang B, Zhang L (1992) *Theory and applications of problem solving*. Elsevier, North-Holland
- Zhang L, Zhang B (2004) The quotient space theory of problem solving. *Fundament inform* 59(2,3):287–298
- Zhang L, Zhang B (2005) Fuzzy reasoning model under quotient space structure. *Inf Sci* 173:353–364
- Zhang L, Zhang B (2007) *The theory and application of problem solving – the theory and application of quotient-space based granular computing*. Tsinghua University Publishing Co. (the second version, in Chinese) Beijing
- Zhu M (2004) Recall, precision and average precision. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo

Non-standard Analysis, an Invitation to

Wei-Zhe Yang
Department of Mathematics, National Taiwan University, Taipei, Taiwan

Article Outline

Glossary
Definition of the Subject
Introduction
Pre-Robinson Infinitesimals
Hyperreals
General Idea of NSA
Loeb Construction
A Future for Non-standard Analysis?
Bibliography

Glossary

Hyperreal A proper extension of the real numbers, containing infinities and infinitesimal numbers, such that a transfer principle allows first-order theorems in the reals to be extended therein.

Infinity An element in an ordered field which is larger than $1 + 1 + 1 + \dots + 1$ for any number of 1's.

Infinitesimal An element in an ordered field whose absolute value is less than $1/n$ for any positive integer n .

Non-archimedean Infinitesimal or infinite values exist in an ordered field.

Transfer principle A rule which transforms assertions about standard sets, mappings, etc., into one about internal sets, mappings. Intuitively, it is an elementary embedding between structures.

Definition of the Subject

Generically speaking, non-standard analysis is any form of mathematics that relies on non-standard models and the transfer principle. However, it is likely safe to say that of the most common meaning for the term “non-standard analysis” is a rigorous formulation of analysis using infinitesimals, as an alternative to the analysis we all know in the $\epsilon - \delta$ formulation. Abraham Robinson is usually regarded as the first person to do this coherently (Robinson 1966).

It is sometimes claimed that non-standard analysis can achieve more than standard models (Bernstein und Robinson 1966). This may not be entirely true, but philosophically, non-standard analysis justifies the standard pedagogical process – infinitesimals in Leibniz notation – in which most people learned calculus, and represents the completion of the a key developmental stage of science (Table 1).

Introduction

Warning 1 *What an (general) undergraduate to a student of mathematics is what a (general) mathematician to a logician.*

This article, written by a (general) mathematician, is aimed at a reader with some mathematical maturity. There are different constructions of non-standard analysis, and we have chosen to start with the “ultrapower” construction of the semantic kind, which is not as difficult to comprehend as the first presentation.

Differential = Infinitesimal Difference?

Suppose a freshman is questioned on the derivative $\frac{dy}{dx}$ for the cubic function $y = x^3$. Then after the difference-quotient is found,

**Non-standard Analysis,
an Invitation to,****Table 1** Notations of this
Chapter

Notation	Meaning
iff	if and only if
$\neg, \wedge, \vee, \Rightarrow$	“negation of”, “and”, “or”, “imply”.
$\forall, \exists$	“for all”, “exists”
$:=$	“defined to be equal to”
$\mathbb{R}, \mathbb{C}, \mathbb{Z}$	set of reals, complex numbers, integers
$\mathbb{N}$	set of natural numbers (zero excluded)
$(a \dots b), [a \dots b]$	open and closed intervals
$(a \dots b], [a \dots b)$	semi-open intervals
$\{\}$	the braces holding a set
$\langle \rangle$	angles holding a sequence
$\cup, \sqcup, \cap, \setminus$	set union, disjoint union, intersection, set difference
$\mathbb{P}(A)$	power set (set of all subsets including A and $\emptyset$)
$\mathbb{R}_\beta^N$	set of all bounded sequences
$\mathbb{R}_c^N$	set of all convergent sequences
$\mathbb{R}_0^N$	set of all sequences converging to zero (=vanishing)
$\mathbb{R}_\kappa^N$	set of all eventually -zero (=nullifying) sequences
$\mathbb{R}^\times$	$=\mathbb{R} \setminus \{0\}$, non-zero reals
$\mathbb{R}_+$	$[0 \dots \infty)$, non -negative real number
$\dot{\mathbb{R}}$	$=\mathbb{R} \sqcup \{-\infty, +\infty\}$, $\mathbb{R}$ with signed infinities
$\bar{\mathbb{R}}$	$=\mathbb{R} \sqcup \{\infty\}$, $\mathbb{R}$ with unsigned infinity
${}^*\mathbb{R}$	(one possible) hyperreal number system
$\mathfrak{U}$	ultra -filter, also and denote (as superscript) “ultrapower construction”
σ	superscript denotes the set of standard elements
$\pi_{\mathfrak{U}}$	denotes the projection for the ultrapower construction $\mathfrak{U}$

$$\frac{\Delta y}{\Delta x} = 3x^2 + 3x(\Delta x) + (\Delta x)^2;$$

he just says “substituting by $\Delta x = 0$ ”, $\frac{dy}{dx} = 3x^2$. And the instructors often have trouble explaining the difference between “ Δx set to 0” and “ $\lim_{\Delta x \rightarrow 0}$ ”.

It is estimated that, among the mathematical-literates (i.e., those who learnt calculus), less than 10% are competent with “ ϵ - δ -logy”. But, of course, they are assured that what they learnt is no nonsense. The calculus we learn is from Leibniz, not Seki Kowa (even if the textbook is in Japanese), and not even Newton, who must have the priority. And the discoveries of calculus and of rigorous analysis (chiefly of Cauchy), are some 150 years apart.

In essence, the success of Leibniz is due to his elegant and mnemonic way of notations, centered

around the notion of differential. When a freshman understands that “the differential of y ”, written as dy , is “the infinitesimal difference of y ”, he deals with calculus problems just like doing routine high school algebra. Indeed calculus as envisaged by Leibniz, is a computation scheme in a number system, denoted by ${}^*\mathbb{R}$ below. This system of “hyperreals” must include, besides the usual real numbers, infinitesimals. Leibniz conceived of the notion of infinitesimals out of necessity, because he was pre-Cauchy. He needed the infinitesimals (as differentials) in the calculus.

But from the very beginning many people, notably Bishop G. Berkeley, “discover impossibilities and contradictions”. Infinitesimals cannot be, simultaneously, zero and non-zero. If they are zero, the division is illegal; if they are not, when $y = x^3$, $\frac{dy}{dx} = 3x^2$ is wrong because $3xdx + (dx)^2$

does not just evaporate away. Of course this is a contradiction that Leibniz, who was a first-rate philosopher-logician, could not fail to discover. Thus the calculus as a science was, for many people before Cauchy, is only recognized de facto and not de jure, just like the nations of Taiwan and Kosovo.

Structure of This Article

In the next section “[Pre-Robinson Infinitesimals](#)”, we will tell you how the people tried to handle infinitesimals before (Robinson 1966). In section “[Hyperreals](#)”, we define the hyperreal field using the ultrapower construction. Section “[General Idea of NSA](#)” builds a small framework of non-standard analysis using hyperreals, and section “[Loeb Construction](#)” introduces the Loeb construction for Wiener processes in NSA.

Pre-Robinson Infinitesimals

Herein we describe the work prior to Robinson (despite say the work of (Schmieden und Laugwitz 1958)), who is credited with being the creator of the field.

Recap: Naive Set Notations

We assume the concept of a set without recourse to the axioms of set theory.

Definition 1 (Binary Boolean Set-Operations)

For two sets A and B , there are these Boolean combinations of them:

$$\begin{aligned} A \cup B &:= \{x : x \in A \text{ or } x \in B\}, \\ A \cap B &:= \{x : x \in A \text{ and } x \in B\}, \\ A \setminus B &:= \{x : x \in A \text{ and } x \notin B\}. \end{aligned}$$

Definition 2 (Function) Let A and B be two non-empty sets. If for any $x \in A$, there is correspondingly assigned one $y \in B$, and written as $f(x)$, then we call this assignment f a function. Its domain is A , codomain B . The notation is $f : A \mapsto B$, or $f : x \in A \mapsto f(x) \in B$. Also, for $C \subset A$, or for $D \subset B$, we write

$$\begin{aligned} f[C] &:= \{f(x) : x \in C\}. \\ f^{-1}[D] &:= \{x \in A : f(x) \in D\}. \end{aligned}$$

If $f[A] = B$, then the function f is surjective. If $f(x) = f(u)$ implies $x = u$, then the function f is injective. When the function is both injective and surjective, it is called bijective. Sometimes the codomain of a function is just understood, but suppressed, and the function is then written as an “indexed family” $\langle f_j : j \in I \rangle$. In this case, we write f_j instead of $f(j)$, and I , the domain of definition, here I , is called the indexing-set, and is often omitted when no confusion is likely. When the indexing set is $\mathbb{N}$ (and in a few other situations) the indexed family is called a **sequence**.

Definition 3 (Cartesian-Product, Couple, Triple, and Sequence) For two non-empty sets A , B , their Cartesian product is defined as

$$A \times B := \{\langle a, b \rangle : a \in A, \text{ and } b \in B\}.$$

Here “a couple”, or “an ordered pair”, is denoted by $\langle a, b \rangle$. An ordered n -tuple is analogously denoted like $\langle x_1, x_2, \dots, x_n \rangle$.

Definition 4 (General Operations Among a Family of Sets) If we have an indexed family of sets $(X_j : j \in I)$, then we define:

$$\begin{aligned} \bigcup_{j \in I} X_j &:= \{x : x \in X_j, \text{ for some } j \in I\}; \\ \bigcap_{j \in I} X_j &:= \{x : x \in X_j, \text{ for each } j \in I\}; \\ \prod_{j \in I} X_j &:= \left\{ g : g \text{ is a function, with } g(j) \in X_j, \right. \\ &\quad \left. \text{for all } j \in I \right\}; \end{aligned}$$

Definition 5 (Quotient-Set and Equivalence Relation) This is a review of the most difficult notion in naive set theory. In studying a non-empty set A , there often occur an “equivalence relation”, denoted by, say, “ $\sim$ ”. This means we write “ $x \sim y$ ” to mean that these two elements x , y of A are “of the same kind”, “similar”, or “equivalent”, referring to something of our concern. If

we collect all similar members into “bunches”, each bunch is called an “equivalence class” (of this relation $\sim$). The set of (bunches=) equivalence-classes is called the quotient set of A with respect to the relation $\sim$, and written as $\underline{A}$. This process of forming the quotient set $\underline{A}$ from the set A , is usually described as “identifying equivalent members”. Formally speaking: a binary relation $\sim$ on A is called an equivalence relation iff it has the following three properties:

- r° (Reflexivity): for any $x \in A$, $x \sim x$.
- s° (Symmetry): when $\{x, y\} \subset A$, and $x \sim y$, then $y \sim x$.
- t° (Transitivity): when $\{x, y, z\} \subset A$, and $x \sim y$, and also $y \sim z$, then $x \sim z$.

For $x \in A$, its equivalence class (modulo $\sim$), is then

$$\pi_\sim(x) := \{y \in A : x \sim y\}.$$

(It is a non-empty subset of A). The quotient set is then:

$$\underline{A} := \{\pi_\sim(x) : x \in A\}.$$

Ordered Field Extension

The relation between the system R of reals and the system ${}^*\mathbb{R}$ of hyper-reals may be announced thus: First of all, ${}^*\mathbb{R}$ should be a proper extension of the ordered-field R . Secondly, one should also be able to extend, from R to ${}^*\mathbb{R}$, the usual functions $\exp$, $\cos$, $\sin$, etc. This is one of the topics of this article, and we explain the meaning of “extension” in this subsection.

Definition 6 (Commutative-Semigroup) Let us consider the arithmetic operation $+$ in $\mathbb{R}$. Though it is a two-variable function, just like, e.g., $f : f(x, y) = \sin(x - y) * (x^3 - y^2)$, it is written “infix”. (So that $x + y$ is “really” $+(x, y)$). There are the noteworthy laws of associativity, and of commutativity:

$$\forall x, \forall y, \forall z, \\ (x + y) + z = (x + (y + z)); \quad x + y = y + x.$$

(We mathematics student read $\forall$ as “for all”). By abstraction, we call a binary operation (= two variable function) on a non-empty set A , a “commutative semigroup structure” on this set. This system, the set A together with the commutative semigroup structure, is called a commutative semigroup. We realize that: in $\mathbb{R}$, multiplication is another commutative semigroup structure.

Definition 7 (Commutative-Group) Now in $\mathbb{R}$, 0 is distinguished by the “unitality law”:

$$\forall x, x + 0 = 0 + x = x.$$

There is then a negation function, mapping x into its negative $-x$, and the reciprocal law holds:

$$x + (-x) = 0; \quad (-x) + x = 0.$$

By abstraction, we call a triple of a two variable function, a distinguished element, and a one-variable function, all on a non-empty set A , satisfying all the laws mentioned above, as a “commutative group structure”. And the set, together with this commutative group structure, is called a commutative group. We realize that: $(\mathbb{R}, (+, 0, -))$ is a commutative group. As to the multiplication, the distinguished unity is 1 , but the reciprocation fails at precisely 0 . If we all the functions are restricted to the zero-deleted set

$$\mathbb{R}^\times := \mathbb{R} \setminus \{0\},$$

then we do have a commutative group structure on it.

Definition 8 (Field) Suppose F is a set of more than one elements. Let the triple $(+, 0, -)$ be a commutative group structure, and $(*, 1, {}^{-1})$ be a commutative group structure on $F^\times = F \setminus \{0\}$. And suppose the distributive law holds:

$$(\forall x, \forall y, \forall z), \quad x * (y + z) = x * y + x * z.$$

Then this couple of triple is called a “field structure” on F . (Of course, there are, altogether, six items in the structure: two distinguished

elements (=“null-variable functions”), two (single variable) functions, two two-variable functions). And the set F , together with this field structure, is called a field.

Definition 9 (Total-Ordering, and Ordered Field) A total-ordering on a set T is a binary relation on T , which satisfies the transitive law and the trichotomy law. For example, in $\mathbb{R}$, the relation “less than” is a total-ordering: the transitive law says that: if $a < b$, and $b < c$, then $a < c$, while the trichotomy law requires that: for any two members a, b , exactly one sentence among the following three is true: $a < b$, $a = b$, $b < a$. An ordered-field structure on a set K is a combination of a field structure, and an total-ordering $<$, with the additional requirement of compatibility law:

$$((0 < x) \wedge (0 < y)) \Rightarrow ((0 < x + y) \wedge (0 < x * y)).$$

$\wedge$ is read “and”, and $\Rightarrow$ is read “imply”. K with this structure, is then an ordered-field. Obviously the set of all rational numbers $\mathbb{Q}$, together with the natural structure, (which is called “inherited from $\mathbb{R}$ ”), is also an ordered-field in this sense.

Definition 10 (Mathematical System, Inheritance and Extension) Returning to the instance of the set $\mathbb{R}$. We may consider the following substructure of five items: the three items of addition commutative-group structure, and the multiplication unital commutative semi-group structure, keeping in mind the distribution law binding them. This is a unital commutative-ring structure. A set K together with a unital commutative-ring structure on it is called a unital commutative-ring. And as is obvious from above discussion, by forgetting the two items in any field structure, we have a unital commutative-ring structure. So the latter is weakened from the former, while the former is strengthened from the latter. Essentially any one particular kind of structure is called a “category”. So within a category one system may be the extension of the other, or the other way round, the latter is a sub-system of the former. For example, the rational ordered field $\mathbb{Q}$ is a sub-system of $\mathbb{R}$, and the real ordered field

$\mathbb{R}$ is an extension of $\mathbb{Q}$. This extension is proper, because $\mathbb{R}$ is strictly bigger: there are irrational numbers. Now for example, the set $\mathbb{Z}$ of all integers is also a unital commutative-ring, when its structure is inherited (with the obvious meaning, or technically “restricted”) from $\mathbb{Q}$. That is, $\mathbb{Z}$ is a sub-unital commutative-ring of $\mathbb{Q}$. But it is not a subsystem of $\mathbb{Q}$ as a field.

Lemma 1 (Quotient Algebra) *Let us mangle usual terminology a little and abbreviate “unital commutative-ring” as “algebra”, just for convenience in this discussion. If A is an algebra, then a subset B is called an ideal, if the following condition is met: for arbitrary $x \in A, y \in B, z \in B$,*

$$y + z \in B, \quad x * y \in B, \\ \text{(which is same as) } y * x \in B.$$

Then we define a (modulo B) relation $\sim_B$ in A , by:

$$(x \sim_B y) \quad \text{iff} \quad (x - y) \in B.$$

And this is an equivalence relation. The equivalence-class of $x \in A$ is then:

$$x + B := \{x + y : y \in B\}.$$

The quotient set $\frac{A}{\sim_B}$ will be written as $A \bmod B$. In it naturally we define the operations by:

$$(x + B) + (y + B) := (x + y) + B, \\ -(x + B) := (-x) + B, \\ (x + B) * (y + B) := (x * y) + B.$$

(And endow it with the natural zero $0 + B$, the unit $1 + B$). The simple result is: these indeed form an algebra structure.

Remark 1 (Other Category) A one variable function is a particular kind of binary relation. And a two-variable function is a particular kind of ternary relation. Indeed a null-variable function (=a distinguished element), is a particular kind of unary relation. Usually structures consisting of these kind of functions, (bound by algebraic

laws), are called “algebraic”. In topology, we study the topological structure, and in measure theory, we need the measurable structure, before strengthening it to the measure structure. And the set $\mathbb{R}$ has on it a natural topological structure and a natural measure structure.

Classical Augmented Infinities

Definition 11 (Leibniz Program) Extend the system $\mathbb{R}$ of reals to the system ${}^*\mathbb{R}$ of hyperreals, so as to contain infinitesimals and also infinities.

This question should be considered in the perspective of “structure of a system” explained in the last subsection. Indeed the NSA of Abraham Robinson is so powerful as to consider the super-structure which is all-encompassing. As far as the size or magnitude (=absolute-value) is concerned, hyperreals should be classified in three categories: infinitesimal, ordinary, and infinite. The reciprocal of an infinitesimal (assumed non-zero), must be infinite (=unlimited), and vice versa. While ordinary hyperreals are neither infinitesimal nor infinite. They are medium-sized, together with their reciprocal. The infinite hyperreals and the infinitesimals (zero excepted), are all non-standard real numbers. The non-zero standard real numbers are all included in the medium-sized hyperreals. Now in calculus, and especially at the beginning pages of complex variables, we sometimes talk about infinities. It turns out that they are standard infinities, and does not belong to the Leibnizian ${}^*\mathbb{R}$ of hyperreals. Let us first try to add one or two points to $\mathbb{R}$, disregarding the infinitesimals for the present. Actually there are precisely three infinities, of two kinds: two signed infinities, and one unsigned infinity.

Definition 12 (Signed Infinities) We denote by $+\infty$ the “positive infinity”, and by $-\infty$ the “negative infinity”. They are added to the totally ordered set $\mathbb{R}$, so as to be, respectively, the maximum and the minimum of this order-extension of $\mathbb{R}$.

$$\dot{\mathbb{R}} = \mathbb{R} \sqcup \{-\infty, +\infty\}.$$

(We use $\sqcup$ here to emphasize that it is a disjoint-union.) So there is an order-isomorphism:

$$x \in \dot{\mathbb{R}} \mapsto \frac{2}{\pi} \arctan(x) \in [-1 \dots 1],$$

(a closed interval)

which is also an homeomorphism (=topological isomorphism). And this compact space $\dot{\mathbb{R}}$ is therefore called the two-point compactification of $\mathbb{R}$.

Definition 13 (Supremum, Infimum) In $\dot{\mathbb{R}}$, any (non-empty) subset A has a supremum, and also an infimum, denoted respectively as $\sup A$, $\inf A$.

We recall that “ b is an upper-bound of A ” means that $\forall x \in A, x \leq b$. And $b = \sup A$ means: b is the least among all upper-bounds of A . In measure theory, most often the values are in the positive half of $\dot{\mathbb{R}}$, that is, the set $[0 \dots +\infty] = \dot{\mathbb{R}}_+$. Finally there is one unsigned infinity, which is the only infinity of the one-point compactification

$$\dot{\mathbb{R}} = \mathbb{R} \sqcup \{\infty\},$$

which is obtained by merging the two signed infinities $\pm\infty$ of $\dot{\mathbb{R}}$ into one point. In $\dot{\mathbb{R}}$, we have, for example:

$$-\infty < -921, -921 < -228, -228 < +\infty;$$

If the ordering were compatible with the merging, then, we would get

$$\infty < -921, -921 < -228, -228 < \infty;$$

and the transitive law is violated! This excludes the possibility of ordering the set $\dot{\mathbb{R}}$. The space $\dot{\mathbb{R}}$ is homeomorphic to the unit circle, e.g. by the (one-dimensional!) stereographic projection

$$x \in \dot{\mathbb{R}} \mapsto (\cos \theta, \sin \theta), \quad \text{where } \theta = 2^* \arctan x - \frac{2}{\pi}.$$

Note: Here we use an unusual notation to distinguish $\infty \in \dot{\mathbb{R}}$ from $+\infty \in \dot{\mathbb{R}}$. Actually these two infinities, one unsigned, the other signed, are never coexistent, and usually they both are

denoted by the same simple ∞ . The deficiency of these systems is that: arithmetic operations involving the infinities are only partially defined in the system, so that there are many undefined operations, so-called “undetermined form”:

$$(+\infty) + (-\infty); \quad (\pm\infty) * 0; \quad (\pm\infty) \div (\pm\infty);$$

Because of these no-definitions, the associative law is greatly compromised. So the use of these infinities is always just symbolic.

Note 1 When $\arctan(x)$ is defined for $x \in \mathbb{R}$, it is valued in $[-\frac{\pi}{2} \dots \frac{\pi}{2}]$, while for $x \in \mathbb{R}$, it is valued in $\mathbb{R} \bmod (\pi)$. Also $\lim_{x \rightarrow 0} (7 - 2x)/x^4 = +\infty \in \mathbb{R}$, while $\lim_{x \rightarrow 0} (7 - 2x)/x^3 = \infty \in \mathbb{R}$. Such distinctions are sometimes useful or necessary.

Non-Archimedean Infinitesimal

Physically speaking, positive real numbers are used for measurements. The possibility of measurement is based on this simple fact: any length λ however long, may be measured against any length ϵ however short. If, for example, $\lambda = 1$ m, and $\epsilon = 3.4$ mm, then $0 < \lambda - n*\epsilon < \epsilon$, with $n = \text{floor}(\frac{\lambda}{\epsilon}) = 294$, so that λ is a little bit more than n units of ϵ . This basic fact is the.

Axiom 1 (of Archimedes) For any $\lambda \in \mathbb{R}_+$, and any $\epsilon \in \mathbb{R}_+$, there is an $n \in \mathbb{N}$, such that $n*\epsilon > \lambda$.

It is precisely this axiom which denies the existence of the infinitesimals: a positive infinitesimal would not be usable as a unit for measurement.

Definition 14 (Infinitesimal) In an ordered field F , an element? is called infinitesimal, if for any $n \in \mathbb{N}$, $n*|\epsilon| < 1$. In other words, no matter how many terms are there, the sum $n*\epsilon = \epsilon + \epsilon + \epsilon + \dots + \epsilon$ (n terms) < 1 .

And in an Archimedean ordered field, the only infinitesimal is zero. When people become acquainted with the concept of limit, they know that “infinitesimal” should not be interpreted statically. Rather, it is the “vanishing behavior” of,

e.g. the function $(\sin^3 x)/x^2$, when $x \rightarrow 0$, that is to be identified with “infinitesimal”. Along this direction there appeared several attempts to find (=construct) the system ${}^*\mathbb{R}$ promised by Leibniz. We now describe a good and very natural try, just years before the discovery of Robinson, by Schmieden–Laugwitz.

Definition 15 (Guiding Idea) A nullifying sequence is considered as zero, a vanishing sequence is considered as an infinitesimal, and a constant sequence is identified with the number itself.

There may be better terminology, but here a sequence $\langle x_n \rangle$ is called nullifying, if after a finite number of terms, the rest of them are all zero. And a sequence $\langle x_n \rangle$ is called vanishing, if it is converging to zero.

Definition 16 (Confusing!) Limit and limiting-point recall the most confusing duo-definitions for a sophomore: “ b is the limit of a sequence $\langle a_n \rangle$ ”, and “ c is a limiting point of a sequence $\langle a_n \rangle$ ”. For any $\epsilon > 0$, there exists an $m \in \mathbb{N}$, such that, $\forall n \geq m$, $|a_n - b| < \epsilon$. For any $\epsilon > 0$, and for any $m \in \mathbb{N}$, there exists an $n \in \mathbb{N}$, such that $n > m$, and $|a_n - c| < \epsilon$. (We support Kelley’s advocacy, replacing “limiting” by “cluster”.)

Definition 17 (Eventual Set, Frequent Set) A subset A of $\mathbb{N}$ is called “eventual”, if it is “cofinite” (=complement of a finite subset). That is: $\mathbb{N} \setminus A$ is finite. It is called “frequent”, if it is infinite. If a property $P(i)$ holds true for i belonging to an eventual set (the mention of which is unnecessary and omitted intentionally), we say $P(i)$ holds true “for eventual i ”. Similar meaning is applied to the phrase “for frequent i ”. With this kinds of abuse of the English language, we may rephrase the definitions above as: For a sequence $a := \langle a_n : n \in \mathbb{N} \rangle$, $b \in \mathbb{R}$ is a limit of a iff, for any $\epsilon > 0$, $a_i \in (b - \epsilon \dots b + \epsilon)$ for eventual i ; c is a limiting-point of a iff, for any $\epsilon > 0$, $a_i \in (c - \epsilon \dots c + \epsilon)$ for frequent i .

Remark 2 See p. 65 in (Kelley 1975), for the definitions of “eventually”, “frequently”, for a

general directed set of indices. Here the (naturally) directed set $\mathbb{N}$ is very special, because it is the least well-ordered infinite set.

Notation 1 (Sequence Spaces) We write $x \in \mathbb{R}^{\mathbb{N}}$ to mean: $x = \langle x_n \rangle$ is a sequence of real numbers. (This is in accordance to our notating the set of mappings). Here are some subsets of $\mathbb{R}^{\mathbb{N}}$, arranged decreasingly:

- $\mathbb{R}_{\beta}^{\mathbb{N}}$ is the set of all bounded sequences;
- $\mathbb{R}_c^{\mathbb{N}}$ is the set of all convergent sequences;
- $\mathbb{R}_0^{\mathbb{N}}$ is the set of all sequences converging to zero (=vanishing);
- $\mathbb{R}_k^{\mathbb{N}}$ is the set of all eventually-zero (=nullifying) sequences.

Definition 18 (Coordinatewise Operations on Sequence Spaces)

$$\begin{aligned}\langle x_n \rangle + \langle y_n \rangle &:= \langle x_n + y_n \rangle; \\ \langle x_n \rangle - \langle y_n \rangle &:= \langle x_n - y_n \rangle; \\ \langle x_n \rangle * \langle y_n \rangle &:= \langle x_n * y_n \rangle.\end{aligned}$$

Lemma 2 The system $\mathbb{R}^{\mathbb{N}}$ is a commutative unital ring, a real vector space, and thence a commutative algebra (with identity). The following inclusions are subalgebra embeddings:

$$\mathbb{R}_c^{\mathbb{N}} \subset \mathbb{R}_{\beta}^{\mathbb{N}} \subset \mathbb{R}^{\mathbb{N}},$$

Now according to the guiding idea, a number $\lambda \in \mathbb{R}$ is identified with the sequence with all terms λ . In this sense $\mathbb{R}_{\beta}^{\mathbb{N}}$, (and a fortiori $\mathbb{R}^{\mathbb{N}}$), is an extension of $\mathbb{R}$, considered as an algebra. But this is not field-extension, because.

Example 1 Let $1_E \in \mathbb{R}_{\beta}^{\mathbb{N}}$ and $1_O \in \mathbb{R}_{\beta}^{\mathbb{N}}$ be the indicator sequences of the sets of even and odd indices, respectively. Then they are complementary idempotents:

$$\begin{aligned}1_E * 1_E &= 1_E; & 1_O * 1_O &= 1_O; & 1_E * 1_O &= 0; \\ 1_E + 1_O &= 1.\end{aligned}$$

Since the two elements are neither 1, nor 0, of the algebra $\mathbb{R}_{\beta}^{\mathbb{N}}$, the latter cannot be a field, because even an eighth grader knows that in a field F , the only idempotent element $x = x^2$ is $x = 0$ (the zero) or $x = 1$ (the identity).

Definition 19 (Extension of Function) Consider these two functions $\cos$ and $\sin$. For a sequence $x = \langle x_n \rangle \in \mathbb{R}^{\mathbb{N}}$, x is a “number” in a generalized sense, and we would like to define $\cos(x)$, $\sin(x)$. If x is a constant sequence: $x_n = a$, $\forall n$, then $\cos(x)$, $\sin(x)$, should be the sequences with constant terms of $\cos(a)$, $\sin(a)$ correspondingly. So we see that a coordinatewise definition is almost forced on us: If $f: \mathbb{R} \mapsto \mathbb{R}$ is everywhere defined, then we define:

$$\bar{f}\langle x_n \rangle := \langle f(x_n) \rangle.$$

So far as it is defined, it works fine. As an example: the addition-formula

$$\begin{aligned}\overline{\sin}(\langle x_n \rangle + \langle y_n \rangle) &= \overline{\sin} \langle x_n \rangle * \overline{\cos} \langle y_n \rangle \\ &\quad + \overline{\sin} \langle y_n \rangle * \overline{\cos} \langle x_n \rangle\end{aligned}$$

would simply mean that, for each index n ,

$$\begin{aligned}\sin(x_n + y_n) &= \sin(x_n) * \cos(y_n) + \sin(y_n) \\ &\quad * \cos(x_n).\end{aligned}$$

Theorem 1 (A Transfer Principle: Restricted-Form) If functions $f, g, \dots$, are all defined on $\mathbb{R}$, and they has a formula connecting them, then this formula still holds for the extended functions on the whole sequence space $\mathbb{R}^{\mathbb{N}}$.

We remark that coordinatewise extension is equally applicable to two-variable function, so long as it is defined everywhere. The Transfer Principle remains valid. And indeed the principle has nothing to do with the differentiability, continuity, etc., of the functions involved. This is pure naive-set theory!

Remark 3 If bounded sequences are called “moderate numbers”, then only bounded

functions will map “moderate number” to “moderate number”.

Definition 20 (Ordering) For two elements $x = \langle x_n \rangle, y = \langle y_n \rangle$ of $\mathbb{R}^{\mathbb{N}}$, we define:

$$(x \leq y), \quad \text{iff} \quad (\forall n, x_n \leq y_n).$$

We write $x < y$, iff $x \leq y$, and $x \neq y$. (This means the following: $x_n \leq y_n$, for all $n \in \mathbb{N}$, but there is an index n , (at least one, and only one is enough!) such that $x_n < y_n$.)

The binary relation $\leq$ in $\mathbb{R}^{\mathbb{N}}$ is called a partial ordering, because it is transitive, and antisymmetric: If $x \leq y$, and also $y \leq x$, then $x = y$. This partial ordering is not a total-ordering, because some pair of members, for example 1_E and 1_O , may be ‘incomparable’, as neither $1_E \geq 1_O$ nor $1_O \geq 1_E$ is true.

Lemma 3 (Nullifying and Vanishing Sequences) *The set $\mathbb{R}_0^{\mathbb{N}}$ is an ideal of $\mathbb{R}^{\mathbb{N}} \supset \mathbb{R}_b^{\mathbb{N}} \supset \mathbb{R}_c^{\mathbb{N}}$. The set $\mathbb{R}_k^{\mathbb{N}}$ is an ideal of $\mathbb{R}_b^{\mathbb{N}} \supset \mathbb{R}_c^{\mathbb{N}}$. Doing modulo $\mathbb{R}_0^{\mathbb{N}}$ means that two sequences with vanishing difference are considered the same. Or, modulo $\mathbb{R}_0^{\mathbb{N}}$ is a magnifying-glass not powerful enough to distinguish a vanishing sequence from zero sequence. And you don't have non-trivial infinitesimals in the system. Also,*

Lemma 4 *The quotient algebra $\mathbb{R}_c^{\mathbb{N}} \text{ mod } \mathbb{R}_0^{\mathbb{N}}$ is (isomorphic to) $\mathbb{R}$.*

The isomorphism identifies any convergent sequence $\langle x_n \rangle$ with its limit $\lim_{n \rightarrow \infty} x_n$. The Lemma says that $\lim_{n \rightarrow \infty} x_n * y_n = \lim_{n \rightarrow \infty} x_n * \lim_{n \rightarrow \infty} y_n$, and the like. (Just the first theorems of our calculus text-book.)

Definition 21 (Quotient by Eventuality) We turn to the quotient algebra $\mathbb{R}_{\mathfrak{E}}^{\mathbb{N}} = \mathbb{R}^{\mathbb{N}} \text{ mod } \mathbb{R}_k^{\mathbb{N}}$, where two sequences $x \in \mathbb{R}^{\mathbb{N}}$ and $y \in \mathbb{R}^{\mathbb{N}}$ are equivalent mod $\mathbb{R}_k^{\mathbb{N}}$, iff $x_n = y_n$ for eventual n . As we denote by $\mathfrak{E}$ the set of all eventual set of indices, we will also write $x \stackrel{\mathfrak{E}}{\sim} y$ or $x =_{\mathfrak{E}} y$ for this equivalence, and the equivalence class of x is written $\pi_{\mathfrak{E}}(x) \in \mathbb{R}_{\mathfrak{E}}^{\mathbb{N}}$. The system $\mathbb{R}_{\mathfrak{E}}^{\mathbb{N}}$ is a gentle modification of $\mathbb{R}^{\mathbb{N}}$. It has essentially the same

good or bad properties of $\mathbb{R}^{\mathbb{N}}$. Because, in essence, the idea of $\text{mod } \mathfrak{E}$ is simply: When studying a sequence, never worry about its few exceptional terms! So we also can extend all the everywhere-defined functions for these “ $\mathfrak{E}$ -numbers”. And the transfer principle still applies! Something better comes out from this modification $\text{mod } \mathfrak{E}$: ordering.

Definition 22 (Ordering) For two elements $x = \langle x_n \rangle, y = \langle y_n \rangle$ of $\mathbb{R}^{\mathbb{N}}$, we write $x \leq_{\mathfrak{E}} y$, and also $\pi_{\mathfrak{E}}(x) \leq \pi_{\mathfrak{E}}(y)$, iff

$$x_n \leq y_n \quad \text{for eventual } n.$$

Naturally we write $x <_{\mathfrak{E}} y$, and also $\pi_{\mathfrak{E}}(x) < \pi_{\mathfrak{E}}(y)$, iff $x \leq_{\mathfrak{E}} y$, and $x \neq_{\mathfrak{E}} y$. This means: $x_n \leq y_n$, for eventual n , but $x_n < y_n$, for frequent n .

Note in the last sentence a subtle difference with the situation in $\mathbb{R}^{\mathbb{N}}$, where “for just one n ” will do. Again this partial ordering is not a total-ordering, as for example $\pi_{\mathfrak{E}}(1_E)$ and $\pi_{\mathfrak{E}}(1_O)$ are still “incomparable”. But the good news is there are “infinitesimals”!

Definition 23 Let $x \in \mathbb{R}^{\mathbb{N}}$. If for all $m \in \mathbb{N}$, $\pi_{\mathfrak{E}}(x) \leq \frac{1}{m}$, then $\pi_{\mathfrak{E}}(x)$ is called an infinitesimal in $\mathbb{R}_{\mathfrak{E}}^{\mathbb{N}}$.

Lemma 5 (Infinitesimals) For $x \in \mathbb{R}^{\mathbb{N}}$, $\pi_{\mathfrak{E}}(x)$ is an infinitesimal $\mathbb{R}_{\mathfrak{E}}^{\mathbb{N}}$, iff $x \in \mathbb{R}_0^{\mathbb{N}}$.

If infinitesimals are just (the equivalence classes of) vanishing sequences, of course sequences “diverging to infinity” are then the infinities!

Hyperreals

The Construction of Ultimate-Reals

Let us review the system $\mathbb{R}_{\mathfrak{E}}^{\mathbb{N}}$, especially the compatibility of the equivalence relation $\stackrel{\mathfrak{E}}{\sim}$ with multiplication:

$$\left((x \stackrel{\mathfrak{E}}{\sim} y) \text{ and } (u \stackrel{\mathfrak{E}}{\sim} v) \right) \Rightarrow (x * u) \stackrel{\mathfrak{E}}{\sim} (y * v).$$

The proof is simple: the conditions $x \mathrel{\mathfrak{E}} y$, $u \mathrel{\mathfrak{E}} v$ mean that $A := \{i : x_i = y_i\} \in \mathfrak{E}$; $B := \{i : u_i = v_i\} \in \mathfrak{E}$; while $A \cap B \subset \{i : x_i^* u_i = y_i^* v_i\}$. And the compatibility condition is guaranteed by the obvious property that:

$$(ii): ((A \in \mathfrak{E}) \& (B \in \mathfrak{E})) \Rightarrow ((A \cap B) \in \mathfrak{E}).$$

On the other hand, the appearance of the zero-divisors, (or equivalently, the nontrivial idempotents), corresponds to the fact that: $(iv^\times)$: there is a set A of indices, such that both A and its complement B is not in $\mathfrak{E}$. E.g., A , and B are even and odd indices of $\mathbb{N}$. With these two remarks in background, we will now make a successful construction of the hyperreals. For the convenience of later possible generalizations, we will write I for $\mathbb{N}$ as the set of indices. Indeed both the well-ordering and the countably infinite cardinality of $\mathbb{N}$ are irrelevant here, and I is required to be an infinite set, just to make the construction interesting. But, for some deeper NSA, an index set of larger cardinality is really needed. So we first form the space $\mathbb{R}^I$, whose elements are real-valued functions on I . Such a function is still called a “sequence”, (even when I is uncountable), and we continue to write $x(j)$ as x_j . As before, binary and unary operations on $\mathbb{R}^I$ are defined coordinatewise. And in particular, it is a commutative algebra with identity.

Definition 24 (Hyperreals) Let $\mathfrak{U}$ be a free ultra-filter on I , i.e., a set of subsets of I , satisfying some conditions to be specified presently. We call two sequences x and $y \in \mathbb{R}^I$ $\mathfrak{U}$ -related, and we write $x \mathrel{\mathfrak{U}} y$, iff:

$$\left\{ j \in I : x_j = y_j \right\} \in \mathfrak{U}.$$

The equivalence class of x is denoted by $\pi_{\mathfrak{U}}(x)$ or $x \bmod \left(\mathrel{\mathfrak{U}} \right)$, and called a hyperreal (along $\mathfrak{U}$).

The quotient set will be denoted simply as ${}^*\mathbb{R}$, $\mathfrak{U}$ being understood. The canonical mapping from $\mathbb{R}^I$ will be denoted by $\pi_{\mathfrak{U}}$.

Definition 25 (Ultra-Filter) A subset $\mathfrak{U}$ of $\mathfrak{P}(I)$ is called a free-ultrafilter on the infinite set I , if the following conditions are fulfilled:

- (i'): if $A \in \mathfrak{U}$ then A is non-empty,
- (i): if $A \in \mathfrak{U}$ then A is not a finite set;
- (ii): if $A \in \mathfrak{U}$ and $B \in \mathfrak{U}$, then $A \cap B \in \mathfrak{U}$;
- (iii): if $A \in \mathfrak{U}$ and $B \supset A$, then $B \in \mathfrak{U}$;
- (iv): if $A \subset I$, then either $A \in \mathfrak{U}$ or $(I \setminus A) \in \mathfrak{U}$.

Remark: We may say that the information about an element $x \in \mathbb{R}^I$ is indexed by I . If $A \in \mathfrak{U}$, then the information about this equivalence class $\pi_{\mathfrak{U}}(x)$ is completely contained in $\langle x_j : j \in A \rangle$, the information about x on that A -indexed part, while the part off A is irrelevant for all(our) purpose. (iii) means: “more than enough (part) is enough”. (ii) means: “if both A part and B part are enough, actually the intersection part is enough”. In mathematical language, if I is any (infinite) set, and $\mathfrak{U} \subset \mathfrak{P}(I)$, $\mathfrak{U}$ is called a filter on I , if it satisfies the condition (i', ii, iii). If the condition (i) also holds, the filter $\mathfrak{U}$ is called free. The dichotomy condition (iv) for an ultra-filter is very strange: “if the information about A part and the information about B part, when gathered jointly, is enough, then actually one or the other part must be enough.

Note 2 (Non-free Ultrafilter and Axiom of Choice) A “non-free ultra”=filter is trivial because it is then the set of all sets of indices containing a particular index i_0 . In that case, $x \mathrel{\mathfrak{U}} y$ would simply mean $x(i_0) = y(i_0)$. In other words: the only relevant information is this i_0 item. The whole indexing of information would be non-sense: there is no need of votes-counting, because the policy is determined by the great-leader uniquely. The existence of a free ultra-filter is guaranteed by the axiom of choice. (Indeed we are told that this is a weaker form of the axiom choice).

From now on, a free ultra-filter $\mathfrak{U}$ is chosen. Again by abuse of English, a set $A \in \mathfrak{U}$ is called an ultimate (set of indices). And “for ultimate j ” means that set of indices j validating the assertion

concerned is in $\mathfrak{U}$. Also, the complement of an ultimate, or by the dichotomy equivalently a non-ultimate, is called an evanescent (or “negligible”) set of indices.

Theorem 2 (Totally Ordered Field) *The quotient system ${}^*\mathbb{R}$ of the algebra $\mathbb{R}^I$ by the equivalence relation $\sim$, is a totally ordered field.*

Proof of the compatibility of the operations and $\sim$: These are all tedious, but they are easy consequences of the filtering condition ii. We will take division as an example. Let $\pi_{\mathfrak{U}}(x) = \pi_{\mathfrak{U}}(u) = \xi \in {}^*\mathbb{R}$, $\pi_{\mathfrak{U}}(y) = \pi_{\mathfrak{U}}(v) = \eta \in {}^*\mathbb{R}$, so that $x \sim u$, $y \sim v$, and we have to show that

$$\left\{ j \in I : \frac{x_j}{y_j} = \frac{u_j}{v_j} \right\} \in \mathfrak{U},$$

and then $\frac{\xi}{\eta} \in {}^*\mathbb{R}$ is unambiguously defined by

$$\frac{\xi}{\eta} := \pi_{\mathfrak{U}} \left\langle \frac{x_j}{y_j} \right\rangle.$$

Of course it is assumed that (in ${}^*\mathbb{R}$), $\eta \neq 0$. That is

$$A := \left\{ j \in I : y_j \neq 0 \right\} \in \mathfrak{U}.$$

Now $\pi_{\mathfrak{U}}(x) = \pi_{\mathfrak{U}}(u)$, $\pi_{\mathfrak{U}}(y) = \pi_{\mathfrak{U}}(v)$, we have:

$$B := \left\{ j \in I : x_j = u_j \right\} \in \mathfrak{U},$$

$$C := \left\{ j \in I : y_j = v_j \right\} \in \mathfrak{U}.$$

And thus $D := A \cap B \cap C \in \mathfrak{U}$. For index $j \in D$, we have:

$$\frac{x_j}{y_j} = \frac{u_j}{v_j},$$

which is what to be shown. Caution: For $j \notin A$, x_j/y_j is not defined! The obvious remedy is thus: a hyperreal $\eta \in {}^*\mathbb{R}$ is represented by a “sequence” $y = (y_j : j \in A)$, which is defined by an ultimate set $A \in \mathfrak{U}$.

Proof of the total ordering: For $x \in \mathbb{R}^I$, $y \in \mathbb{R}^I$, consider the sets $A := \{j : x_j \geq y_j\}$, $B := \{j : y_j \geq x_j\}$. Now $A \cup B = I \in \mathfrak{U}$, so either $A \in \mathfrak{U}$, or $B \in \mathfrak{U}$, by

dichotomy. Thus $\pi_{\mathfrak{U}}(x) \geq \pi_{\mathfrak{U}}(y)$ or $\pi_{\mathfrak{U}}(y) \geq \pi_{\mathfrak{U}}(x)$. Needless to say: when $A \in \mathfrak{U}$, and $B \in \mathfrak{U}$, then $A \cap B = \{j : x_j = y_j\} \in \mathfrak{U}$, thus $x \sim y$. Note that in the proof above for division, we actually proved the multiplicative invertibility for $\eta \neq 0$, therefore we have shown that in

$${}^*\mathbb{R}$$

there is no non-trivial idempotent. as there is no zero-divisors.

Definition 26 (Infinitesimals and Infinities)

Since $\mathbb{R}$ is embedded canonically in the totally ordered field ${}^*\mathbb{R}$, a hyperreal a is called infinitesimal iff: for any $n \in \mathbb{N}$, $|a| < \frac{1}{n}$. The set of all infinitesimals is denoted by $\mathbb{I}_0$, while the subset of all non-zero infinitesimals is denoted by $\mathbb{I}$.

The set $\mathbb{I}_0$ is a real linear subspace of ${}^*\mathbb{R}$. Note that the only standard member therein is zero. So $\mathbb{I} \cap \mathbb{R} = \emptyset$.

Definition 27 (Monad) Any two hyperreals with infinitesimal difference are called infinitely close. This is an equivalence relation of “infinitely close”, written $\approx$, whose equivalence classes are called “monads”, after Leibniz.

Definition 28 (Limited and Unlimited) A hyperreal $\alpha \in {}^*\mathbb{R}$ is called limited (=“finite”) if, for some $n \in \mathbb{N}$, $-n \leq \alpha \leq n$. In this case, there is uniquely one real number $a = {}^\circ\alpha$ infinitely close to α . Then $\alpha = a + (\alpha - a)$, a is the standard part of α , and $\alpha - a$ is the infinitesimal part of α . On the other hand, a hyperreal α is called limited (“infinite”) iff: for any $n \in \mathbb{N}$, $|\alpha| > n$.

An equivalent definition of this infinity is that: $\frac{1}{\alpha} \in \mathbb{I}$. The set of all limited hyperreals is a sub-algebra (with identity) of ${}^*\mathbb{R}$, with the set $\mathbb{I}_0$ of infinitesimals as an ideal. The quotient algebra is isomorphic with the real field $\mathbb{R}$, so the set of all limited hyperreals is indeed

$$\mathbb{R} + \mathbb{I}_0 = \{c + \zeta : c \in \mathbb{R}, \zeta \in \mathbb{I}_0\}.$$

Note: Keeping $I = \mathbb{N}$, and O, E the subsets of odd and even indices respectively, one of the two

sequences $1_O, 1_E$, is (equivalent to) the unit ($=1$) of ${}^*\mathbb{R}$, while the other is the zero ($=0$) of ${}^*\mathbb{R}$. (They are idempotents complementing each other). The choice is made by the ultrafilter $\mathcal{U}$. Or, we can choose one or the other from the two sets O, E , to be a member of $\mathcal{U}$. The axiom of choice allow us to do this before the final set $\mathcal{U}$ is ultimately found.

Example 2 (A Monad Is at Least a Continuum)

When $I = \mathbb{N}$, the hyperreal $\varepsilon = \pi_{\mathcal{U}}\langle \frac{1}{n} : n \in \mathbb{N} \rangle$ is infinitesimal. Then so is ε^3 , and indeed for any positive real number $r > 0$, ε^r is infinitesimal. These ε^r are inversely ordered as r . This is easily seen from the representing sequences $\langle n^{-r} : n \in \mathbb{N} \rangle$. So the monad of infinitesimals is at least a continuum, cardinally speaking.

Example 3 (A Smaller Infinitesimal) Suppose real numbers $r > 1, s > 1$ are fixed. Then for eventual $n \in \mathbb{N} = I$,

$$s^{-n} < \frac{1}{n^r};$$

A fortiori, this is true for ultimate n ; therefore

$$v_s = \pi_{\mathcal{U}}\langle s^{-n} : n \in \mathbb{N} \rangle < \varepsilon^r.$$

To an atheist-Buddhist, an ultrafilter $\mathcal{U}$ is a Buddha's magnifying glasses, viewing through which a tiny monad is bigger than a solar system.

Star-Extensions

Now we come to the central issue of function-extension and the intimately related transfer principle. For an everywhere-defined function f , the coordinatewise definition still works fine, and we surely have an extension $\bar{f}$ of f to the hyperreals. (The same proof works. Or rather, the assertion here is then a corollary of the former principle.) For a function like csc , which is undefined at $n\pi \in \pi^*\mathbb{Z}$, $\bar{\text{csc}}$ as definable at $x = \langle x_n \rangle$, is

$$\bar{\text{csc}}(x) := \langle \text{csc}(x_n) \rangle \in \mathbb{R}^{\mathbb{N}}.$$

In $\mathbb{R}^{\mathbb{N}}$, the definition breaks down, if only one term $x_n \in \pi^*\mathbb{Z}$. How about the situation in $\mathbb{R}_{\varepsilon}^{\mathbb{N}}$? Write temporarily the domain of definition of csc

as $A := \text{Dom}(\text{csc}) = \mathbb{R} \setminus \pi^*\mathbb{Z}$, then if for eventual $n, x_n \in B$, the definition goes through! And for such an $x \in \mathbb{R}^{\mathbb{N}}$, $\bar{\text{csc}}(\pi_{\mathcal{U}}(x)) \in \mathbb{R}_{\varepsilon}^{\mathbb{N}}$ is defined. In complete analogy, for an $x \in \mathbb{R}_{\varepsilon}^{\mathbb{N}}$, if only: “for ultimate $n, x_n \in A$ ”, $\pi_{\mathcal{U}}\langle \text{csc}(x_n) \rangle \in {}^*\mathbb{R}$ is well-defined!

Definition 29 (Star-Extension of a Set) For a non-empty set $A \subset \mathbb{R}$, the set *A is the set of all $\pi_{\mathcal{U}}(x)$, where $x = \langle x_j \rangle$ is a “sequence” such that, for ultimate $j, x_j \in A$.

Definition 30 (Extension of a Mapping) For any mapping $f: A \mapsto B$, and $x = \langle x_j \rangle \in A^I$, we have the “sequence” $y = f \circ x = \langle f(x_j) \rangle \in B^I$, whose equivalence class $\eta = \pi_{\mathcal{U}}(y) \in {}^*B$ is unambiguously determined by $\xi = \pi_{\mathcal{U}}(x) \in {}^*A$, and is independent of the representative $x \in A^I$. We then write $\eta = {}^*f(\xi)$. This mapping ${}^*f: {}^*A \mapsto {}^*B$ is called the star-extension of f . (It is surely an extension.)

It is of utmost importance to note here that these definitions of “star-extensions” are strictly notions of naive set theory. Of course A, B can be any two non-empty sets, and f any function from A to B .

Theorem 3 (Transfer Principle) *All the identities for the standard functions hold true for their star-extensions, simply because they hold true for the original functions.*

Example 4 (Hyper Integers) A hyperinteger $\pi_{\mathcal{U}}(x) \in {}^*\mathbb{Z}$ is the equivalence class of a sequence of integers $x \in \mathbb{Z}^{\mathbb{N}}$. If “integer” is changed to “even integer”, then we get even hyperinteger $\pi_{\mathcal{U}}(x) \in {}^*(2^*\mathbb{Z})$, where $2^*\mathbb{Z} := \{2^n : n \in \mathbb{Z}\}$ is the set of even integers. And there are odd hyperintegers, too. All the hyperintegers may be classified as even or odd, because of the dichotomy property of the ultrafilter $\mathcal{U}$.

Example 5 (Cartesian Product and Two Variable Function) For two non-empty sets A, B , by canonical identification,

$$^*(A \times B) = ^*A \times ^*B.$$

So for any mapping $g: A \times B \mapsto C$, the star-extension of g may be defined: $^*g: ^*A \times ^*B \mapsto ^*C$.

What this says is easy but tedious to explain. First, we know the identification:

$$(A \times B)^I = A^I \times B^I.$$

Because for $x \in A^I, u \in B^I$, the couple $\langle x, u \rangle \in A^I \times B^I$ corresponds bijectively to an element $x \otimes u$ of $(A \times B)^I$, by:

$$x \otimes u : i \in I \mapsto \langle x_i, u_i \rangle \in A \times B;$$

And the equivalence class of $x \otimes u$ depends on the equivalence classes $\pi_{\mathcal{U}}(x)$ and $\pi_{\mathcal{U}}(u)$ only. In this way we are sure that $\pi_{\mathcal{U}}\langle g(x_j, y_j) \rangle$ is unambiguously determined by $\pi_{\mathcal{U}}(x)$ and $\pi_{\mathcal{U}}(u)$ only, and is independent of the representative $x \in A^I, y \in B^I$. Actually we did give a proof for $g =$ division, $A = \mathbb{R} = C, B = \mathbb{R}^\times = \mathbb{R} \setminus \{0\}$. The proof is good for any function of any (finite) number of variables. Some examples of Transfer are:

- $^*\text{abs}(a) = |a| \in ^*\mathbb{R}$ is meaningful for $a \in ^*\mathbb{R}$. And indeed $|a| = -a$, if $a < 0$.
- $(-1)^\mu$ is meaningful for $\mu \in ^*\mathbb{Z}$. And indeed it is 1 for μ an even hyperinteger, it is -1 for μ an odd hyperinteger.
- The star-extensions $^*\sin, ^*\cos$ of the sine, cosine functions are defined on $^*\mathbb{R}$ and valued in the star-extended unit closed interval $^*[0 \dots 1]$.

Note: From now on, the star-extensions of such classical functions like $\sin, \cos, \dots$, will not have the star displayed.

NS View of Continuous Functions

The introduction of hyperreals gives us a different standpoint to view many familiar concepts.

Remark 4 (Extended Version of a Sequence)

Suppose $s = \langle s_n : n \in \mathbb{N} \rangle$ is a sequence of real

numbers. This is a function from $\mathbb{N}$ to $\mathbb{R}$. Therefore, it may be star-extended to a function

$$^*s : m \in ^*\mathbb{N} \mapsto s_m \in ^*\mathbb{R}.$$

We call this latter the extended version of the original sequence. Indeed for any unlimited integer $v = \pi_{\mathcal{U}}\langle n_j : j \in \mathbb{N} \rangle$, we may define

$$s_v := \pi_{\mathcal{U}}\langle s_{n_j} : j \in \mathbb{N} \rangle \in ^*\mathbb{R}.$$

It is obvious that the convergence of the sequence s

$$\lim_{n \rightarrow \infty} s_n = b,$$

is equivalent to:

for all unlimited integer $v \in ^*\mathbb{N} \setminus \mathbb{N}$, $s_v \approx b$.

As an example, when $r \in \mathbb{R}, y \in \mathbb{R}$, and

$$s_n := \cos^m\left(\frac{y}{\sqrt{n}}\right), \quad \text{with } m = \text{floor}(n * r),$$

standard calculus gives us

$$\lim_{n \rightarrow \infty} s_n = e^{-\frac{y^2 r}{2}}.$$

We will talk about a function $f: \mathbb{R} \mapsto \mathbb{R}$, though usually the domain of definition is only required to be an open interval.

Theorem 4 A function f is bounded around a point $a \in \mathbb{R}$, iff *f maps the monad of a into finite hyperreals.

Proof If f maps a neighborhood $(a - e \dots a + e)$ of a into $[-M \dots M]$, where $e \in \mathbb{R}$ is positive, then for $\xi = \pi_{\mathcal{U}}\langle x_j \rangle \approx a$, we have $a - e < x_j < a + e$ for ultimate j , and thence $f(x_j) \in [-M \dots M]$. Accordingly:

$$-M^* \leq ^*f(\xi)^* \leq M^*.$$

On the other hand, if f is unbounded around a , then for any $j \in I = \mathbb{N}$, there is x_j (in the domain of definition of f), with $|x_j - a| < \frac{1}{j}$, and $f(x_j) > j$. Thus $\xi := \pi_{\mathcal{U}}\langle x_j \rangle \approx a$, but $|*f(\xi)|$ is infinite. $\square$

Theorem 5 Let $A \subset \mathbb{R}$ be an interval, $b \in A$, and $f: A \Rightarrow \mathbb{R}$. Then f is continuous at b , iff: for all $\xi \in {}^*A$,

$$\xi \approx b \Rightarrow *f(\xi) \approx f(b).$$

In other words, f is continuous at b , iff $*f$ maps the monad $a + \mathbb{I}_0$ into the monad $f(a) + \mathbb{I}_0$.

Proof By the standard definition, when f is continuous at b , and $\epsilon > 0$ is given, there is a $\delta > 0$, such that, for $|x - b| < \delta$, (and $x \in A$), $|f(x) - f(b)| < \epsilon$. Now $\xi = \pi_{\mathcal{U}}\langle x_j \rangle$, and $\xi - b \in \mathbb{I}_0$, so for ultimate j , $|x_j - b| < \delta$. Thus $|f(x_j) - f(b)| < \epsilon$, for ultimate j . This means $*f(\xi) - f(b) \in \mathbb{I}_0$, or, $*f(\xi) \approx f(b)$. Suppose f is not continuous at b , then there is $\epsilon > 0$, such that for whatever $j \in \mathbb{N}$, there is $x_j \in A$, with $|x_j - b| < \frac{1}{j}$ while $|f(x_j) - f(b)| > \epsilon$. Then with $\xi = \pi_{\mathcal{U}}\langle x_j \rangle \in {}^*A$, $\xi \approx b$, while $|*f(\xi) - f(b)| > \epsilon$. $\square$

Corollary 1 If f is continuous at c , it is bounded around c .

Theorem 6 Let $A \subset \mathbb{R}$ be a compact (=closed and bounded) interval, and $f: A \mapsto \mathbb{R}$ is continuous. Then f is bounded.

Proof If f is unbounded, then for $j \in \mathbb{N}$, there is $x_j \in A$, with $|f(x_j)| > j$. And we find $\xi = \pi_{\mathcal{U}}\langle x_j \rangle$, with $*f(\xi)$ infinite. But $\xi \in {}^*A$ is a limited hyperreal, and its standard part $b = {}^\circ\xi \in A$, $b \approx \xi$. By the continuity criterion, $*f(\xi) \approx f(b) \in \mathbb{R}$. This is a contradiction.

Theorem 7 Let $A \subset \mathbb{R}$ be an interval, and $f: A \mapsto \mathbb{R}$. Then f is uniformly continuous, iff: for all $\xi \in {}^*A$, $\eta \in {}^*A$,

$$\xi \approx \eta \Rightarrow *f(\xi) \approx *f(\eta).$$

Proof By the standard definition, when f is uniformly continuous on A , and $\epsilon > 0$ is given, there

is a $\delta > 0$, such that, for $|x - y| < \delta$, (and $x \in A$, $y \in A$), $|f(x) - f(y)| < \epsilon$. Now $\xi = \pi_{\mathcal{U}}\langle x_j \rangle \in {}^*A$, $\eta = \pi_{\mathcal{U}}\langle y_j \rangle \in {}^*A$, and $\xi - \eta \in \mathbb{I}_0$, so for ultimately many j , $|x_j - y_j| < \delta$. Thus $|f(x_j) - f(y_j)| < \epsilon$, for ultimately many j . This means $*f(\xi) - *f(\eta) \in \mathbb{I}_0$, or, $*f(\xi) \approx *f(\eta)$. Suppose f is not uniformly continuous, then there is $\epsilon > 0$, such that for whatever $j \in \mathbb{N}$, there is a pair $x_j \in A$, $y_j \in A$, with $|x_j - y_j| < \frac{1}{j}$, while $|f(x_j) - f(y_j)| > \epsilon$. Then with $\xi = \pi_{\mathcal{U}}\langle x_j : j \rangle \in {}^*A$, $\eta = \pi_{\mathcal{U}}\langle y_j : j \in I \rangle \in {}^*A$, $\xi \approx \eta$, while $|*f(\xi) - *f(\eta)| > \epsilon$. Contradiction. One of the merits of NSA is that: a standard notion may be interpreted in a nonstandard way, which is pedagogically advantageous. There will be, as examples, a couple of theorems easily proved by the NS criterion just given. But first we introduce some notions very useful later on. $\square$

Definition 31 (Hyperfinite Set) Let $T_j \subset A$ for each $j \in I$. Then the sequence defines an “internal” subset $\mathcal{T} := \pi_{\mathcal{U}}(T)$ of $*A$:

$$\pi_{\mathcal{U}}(T) := \{ \pi_{\mathcal{U}}(x) : x_j \in T_j \text{ for ultimate } j \} \subset {}^*A.$$

If for ultimate j , the set T_j is finite: $\text{card}(T_j) < \infty$, $\mathcal{T}$ is called a hyperfinite set with internal cardinal $\pi_{\mathcal{U}}\langle \text{card}(T_j) \rangle \in {}^*\mathbb{N}$.

Example 6 (Hyperfinite Equi-Partition) Consider the hyperfinite set $\mathcal{T}$ of equi-partition points with cardinal $1 + \mathcal{N}$:

$$\mathcal{N} := \pi_{\mathcal{U}}\langle N_j \rangle \in \mathbb{N}^I; (N_j < N_{j+1} \rightarrow \infty)$$

$$T_j := \left\{ \frac{k}{N_j} : k \in \mathbb{Z}, 0 \leq k \leq N_j \right\};$$

$$\mathcal{T} := \pi_{\mathcal{U}}\langle T_j \rangle.$$

Three convenient choices are $N_j = j!$ and $N_j = 2^j$, and $N_j = 3^j$. The corresponding partition set $\mathcal{T}$ is then called, respectively, the partition by factorial, or binary, or ternary fractions.

Lemma 6 (Density of Equi-Partition Set) The countable set

$$T_\infty = \cup_j T_j$$

is dense in the unit interval $\mathbb{U}$: For any number $c \in \mathbb{U}$, there is a hyperreal $\zeta \in {}^*T$, $\zeta \approx c$.

Proof There are two convenient choices: the left- and the right-approximation:

$$\zeta = \text{floor}_T(c) = \pi_{\mathbb{U}} \left\langle \frac{\text{floor}(c^* N_j)}{N_j} : j \in I \right\rangle;$$

$$\text{or } \zeta = \text{ceil}_T(c) = \pi_{\mathbb{U}} \left\langle \frac{\text{ceil}(c^* N_j)}{N_j} : j \in I \right\rangle.$$

Note that, for any $x \in \mathbb{U}$,

$$\text{st}(\text{floor}_T(x)) = x, \quad \text{st}(\text{ceil}_T(x)) = x.$$

Both floor_T and ceil_T are (Skolem?) “inverse-function” of the standard-part mapping st from $\mathbb{U}$ onto T . Indeed for the binary-rational partition $T_j = 2^j$, or the ternary-rational partition $T_j = 3^j$, the floor-approximation gives the binary or ternary expansion respectively.

Theorem 8 Let $A \subset \mathbb{R}$ be a compact interval, and $f: A \Rightarrow \mathbb{R}$ is continuous. Then f has a maximum and minimum.

Proof Now (by simple scaling and translations) we may and will assume $A = \mathbb{U} = [0 \dots 1]$ the unit interval. Take a hyperinteger $\mathcal{N} := \pi_{\mathbb{U}} \langle N_j \rangle$, and then the equi-partition $T = \pi_{\mathbb{U}}(T)$ into $\mathcal{N}$ parts. Here $T_j := \{k/N_j : k \in \mathbb{Z}, 0 \leq k \leq N_j\}$. Now for each j , the maximum value of

$$f[T_j] := \{f(x) : x \in T_j\}$$

is unique, though the maximum point x_j may be non-unique:

$$f(x_j) \geq f(x), \quad \forall x \in T_j;$$

We just choose (say) the least one such maximum point x_j . And define

$$\xi := \pi_{\mathbb{U}}(x) \in {}^*A; \quad x_0 := {}^\circ \xi; \quad x_0 \in A.$$

Then for any $c \in A$, and ζ as before,

$${}^*f(\zeta) \leq {}^*f(\xi); \quad \zeta \approx c;$$

$$x_0 \approx \xi; \quad {}^*f(x_0) \approx {}^*f(\xi); \quad {}^*f(c) \approx {}^*f(\xi);$$

We see that $f(c) \leq f(x_0)$, $\forall c \in A$, i.e., f attains its maximum-value at x_0 . $\square$

Theorem 9 (Intermediate-Value) Let $f: A \mapsto \mathbb{R}$ be continuous, $v \in \mathbb{R}$, $A = [a \dots b]$, and $(v - f(a))(v - f(b)) < 0$. Then there is $u \in (a \dots b)$, such that $f(u) = v$. $\square$

Proof We may and will assume $A = \mathbb{U}$, $a = 0$, $b = 1$, $f(0) < 0 < f(1)$. As before, we take a hyper-finite $\mathcal{N}$ equi-partition $T = \pi_{\mathbb{U}} \langle T_j \rangle$, $\mathcal{N} = \langle N_j \rangle$. Find the least $x_j \in T_j$, such that $f(x_j) \geq 0$. Because $0 \in T_j$, $1 \in T_j$, and $f(0) < 0 < f(1)$, we know $f(x_j) \geq 0$, $f(x_j - 1/N_j) < 0$. With $\xi := \pi_{\mathbb{U}} \langle x_j \rangle \in {}^*A$, $u := {}^\circ \xi$, then $\epsilon := \pi_{\mathbb{U}} \langle 1/N_j \rangle$, we have:

$${}^*f(\xi) \geq 0, \quad {}^*f(\xi - \epsilon) < 0.$$

But $\xi \approx u$; $\xi - \epsilon \approx u$; and the continuity of f at u implies:

$${}^*f(\xi) - f(u) \in \mathbb{I}_0, \quad {}^*f(\xi - \epsilon) - f(u) \in \mathbb{I}_0.$$

Therefore:

$$\begin{aligned} 0 &\leq {}^*f(\xi) < {}^*f(\xi) - {}^*f(\xi - \epsilon) \\ &= ({}^*f(\xi) - f(u)) - ({}^*f(\xi - \epsilon) - f(u)) \in \mathbb{I}_0, \end{aligned}$$

resulting in ${}^*f(\xi) \in \mathbb{I}_0$. Or ${}^\circ({}^*f(\xi)) = 0$, which means $f(u) = 0$.

Theorem 10 Let $A \subset \mathbb{R}$ be a compact interval, and $f: A \mapsto \mathbb{R}$ be continuous. Then f is uniformly continuous.

Proof Suppose $\xi \in {}^*A$, $\eta \in {}^*A$, and $\xi \approx \eta$. Now let $c = {}^\circ(\xi) \in A$. It follows that both $\xi \approx c$ and $\eta \approx c$ hold true. The continuity of f implies that both

${}^*f(\xi) \approx f(c)$ and ${}^*f(\eta) \approx f(c)$ hold. Therefore ${}^*f(\xi) \approx {}^*f(\eta)$. $\square$

NS Calculus

According to the standard definition:

$$f'(a) = \lim_{\Delta x \rightarrow 0} \frac{f(a + \Delta x) - f(a)}{\Delta x};$$

the differentiability of f at a implies the continuity of f at a , which means that *f maps the monad $a + \mathbb{I}_0$ of a into the monad $f(a) + \mathbb{I}_0$ of $f(a)$.

Definition 32 (Difference) If f is a real-valued function defined around a , then the “discern” of f at a is the (non-standard) function

$$\Delta_a f : dx \in \mathbb{I}_0 \mapsto {}^*f(a + dx) - f(a) \in {}^*\mathbb{R}.$$

Of course we know that the continuity of f at a means that the range of $\Delta_a f$ is included in $\mathbb{I}_0$. So we are led to consider (non-standard) functions from $\mathbb{I}_0$ to $\mathbb{I}_0$. From now on, $dx, dy, \dots$ will be variables whose values are infinitesimals. (So they live in $\mathbb{I}_0$.)

Definition 33 (Tangential and Tangent) If $\Phi : \mathbb{I}_0 \mapsto \mathbb{I}_0, \Psi : \mathbb{I}_0 \mapsto \mathbb{I}_0$, and for any nonzero infinitesimal $\epsilon \in \mathbb{I}$,

$$\frac{\Phi(\epsilon) - \Psi(\epsilon)}{\epsilon} \approx 0,$$

they are called tangential to each other. For any number $c \in \mathbb{R}$, the associated tangent is the (non-standard) function:

$$dx \in \mathbb{I}_0 \mapsto c^* dx \in \mathbb{I}_0.$$

Theorem 11 (Differentiability) A function f defined on an open interval U is differentiable at $a \in U$, and there with derivative $f'(a) = c$, iff: $\Delta_a f$ is tangential to the tangent associated with c .

Proof If $f'(a) = c$, then for any $k \in \mathbb{N}$, there is $e_k \in \mathbb{R}$ with $e_k > 0$, such that whenever $|u| < e_k$, $|(f(a + u) - f(a))/u - c| < \frac{1}{k}$. Now fix $k \in \mathbb{I}$. Thus to any (non-zero) $dx = \pi_{\mathbb{U}}\langle u_j \rangle \in \mathbb{I}$, for

ultimate j , $0 < |u_j| < e_k$, resulting in $|(f(a + u_j) - f(a))/u_j - c| < \frac{1}{k}$. So

$$\left| \frac{\Delta_a f(dx) - c^* dx}{dx} \right| = \left| \frac{{}^*f(a + dx) - f(a)}{dx} - c \right| < \frac{1}{k}.$$

This being true for all $k \in \mathbb{N}$, we have: for any $dx \in \mathbb{I}$,

$$\left| \frac{\Delta_a f(dx) - c^* dx}{dx} \right| \in \mathbb{I}_0;$$

which means: $\Delta_a f$ is tangential to the tangent associated with c . Conversely, if f is not differentiable at a , one can show that the nonstandard condition of differentiability can not be satisfied. Indeed then the following four limits of the difference quotient $\frac{f(a+u)-f(a)}{u}$ are not all finite and coincident:

$$\limsup_{u \downarrow 0}, \quad \liminf_{u \downarrow 0}, \quad \limsup_{u \uparrow 0}, \quad \liminf_{u \uparrow 0}.$$

For example, a case may be

$$\begin{aligned} b &:= \limsup_{u \downarrow 0} \frac{f(a + u) - f(a)}{u} \\ &> e := \liminf_{u \downarrow 0} \frac{f(a + u) - f(a)}{u}. \end{aligned}$$

In that case, there are positive sequences $u \in \mathbb{R}^{\mathbb{N}}, v \in \mathbb{R}^{\mathbb{N}}$, decreasing to 0, such that

$$\begin{aligned} b &:= \lim_{j \rightarrow \infty} \frac{f(a + u_j) - f(a)}{u_j}, \\ e &:= \lim_{j \rightarrow \infty} \frac{f(a + v_j) - f(a)}{v_j}. \end{aligned}$$

Taking $\xi = \pi_{\mathbb{U}}\langle u_j \rangle \in \mathbb{I}$, and $\eta = \pi_{\mathbb{U}}\langle v_j \rangle \in \mathbb{I}$, we see that: the nonstandard condition of differentiability can not be satisfied simultaneously for $dx = \xi$ and for $dx = \eta$, whatever the choice of c .

Theorem 12 (Chain-Rule) If a function f is differentiable at a , and a function g is differentiable

at $b = f(a)$, then the composite function $h := g \circ f$ is differentiable at a , with derivative $h'(a) = g'(b) * f'(a)$.

Proof For any $\xi \in \mathbb{I}$, $\eta \in \mathbb{I}$, we have:

$$\frac{*f(a + \xi) - f(a)}{\xi} \approx f'(a) ;$$

$$\frac{*g(b + \eta) - g(b)}{\eta} \approx g'(b) .$$

Let now $\eta = *f(a + \xi) - f(a)$. Then:

$$\begin{aligned} \frac{*h(a + \xi) - h(a)}{\xi} &= \frac{*g(*f(a + \xi)) - g(f(a))}{\xi} \\ &= \frac{*g(b + \eta) - g(b)}{\eta} \cdot \frac{\eta}{\xi} \approx g'(b) * f'(a) . \end{aligned}$$

Here it is assumed that $\eta \neq 0$ so as to allow the division by η . But when $\eta = 0$, then necessarily $f'(a) = 0$, and the proposed identity is trivial. $\square$

Theorem 13 (Leibniz' Product Rule) *If two functions f and g are differentiable at a , then so is their product h , $h(x) = f(x) * g(x)$. And.*

$$h'(a) = f'(a) * g(a) + f(a) * g'(a).$$

Proof The non-standard calculations are essentially the same as the standard ones. $\square$

Lemma 7 *If f has a local maximum at $x = a$, then.*

$$\forall dx \in \mathbb{I}, \quad \Delta_a f(dx) \leq 0.$$

Theorem 14 ("Fermat") *On an open interval, if a function f has a maximum at a , and there differentiable, then $f'(a) = 0$.*

Proof Recall the signum function $\text{sign}(x) = \frac{x}{|x|}$ for $x \neq 0$. We will not put the star on its star-extension. For $dx \in \mathbb{I}$,

$$\frac{\Delta_a f(dx) - f'(a)dx}{dx} \in \mathbb{I};$$

So if $f'(a) \neq 0$, we take $dx \in \mathbb{I}$ to have the same signum as $f'(a)$, then there is a contradiction:

$$\text{sign}(\Delta_a f(dx)) = \text{sign}(f'(a) * dx) > 0.$$

$\square$

Conventional deductions then give, as corollaries,

Theorem 15 ("Rolle") *Let a function f is continuous on a closed interval $[a \dots b]$, and differentiable on the open interval $(a \dots b)$. If moreover $f(a) = f(b)$, then there exists a zero of the derived function f' in the open interval. Generally, there exists ξ in the open interval $(a \dots b)$, such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

has a maximum at a , and there differentiable, then $f'(a) = 0$.

Remark 5 For a function f defined on a closed interval, one-sided derivative may still be defined at the end-points naturally. Nelson mentioned that a stronger notion is in some way more convenient than differentiability. f is called derivable at a point a , iff there is a derivative $f'(a) = c$ such that:

$$\begin{aligned} &(\text{as } x_1 \rightarrow a, \ x_2 \rightarrow a, \ x_1 \neq x_2), \\ &\lim \left(\frac{f(x_2) - f(x_1)}{x_2 - x_1} - f'(a) \right) = 0. \end{aligned}$$

The NS formulation is:

$$\begin{aligned} &(\epsilon_1 \in \mathbb{I}, \ \epsilon_2 \in \mathbb{I}, \ \epsilon_1 \neq \epsilon_2) \\ &\Rightarrow \frac{*f(a + \epsilon_2) - *f(a + \epsilon_1)}{\epsilon_2 - \epsilon_1} \approx f'(a). \end{aligned}$$

Of course all the differentiation rules are valid for this stronger-sense derivative.

Theorem 16 *A function f is derivable on an interval, (i.e., at every point of the interval), iff it is differentiable with continuous derivative everywhere on the interval.*

The function $f(x) = x^2 \cos(\frac{1}{x})$ (extended by $f(0) = 0$), is differentiable on $\mathbb{R}$, but $f'(x) = 2x \cos(\frac{1}{x}) + \sin(\frac{1}{x})$ (extended by $f'(0) = 0$), is not continuous at $x = 0$. Indeed it is not derivable there.

NS Integral Calculus

Let $N := (N_j : j \in I) \in \mathbb{N}^I$ be a strictly increasing sequence of natural numbers, and for each $j \in I$, define

$$T_j := \left\{ \frac{k}{N_j} : k \in \mathbb{Z}, 0 \leq k \leq N_j \right\}.$$

The unit interval $\mathbb{U}$ is correspondingly partitioned into N_j subintervals $[(k-1)/N_j \dots k/N_j]$, $k = 1, 2, \dots, N_j$. Suppose f is a function on the unit interval $\mathbb{U}$. The left-Riemann sum of f , relative to this partition is

$$S_j := \frac{1}{N_j} \sum (f(x) : x \in T'_j), \quad T'_j := T_j \setminus \{1\}.$$

Assuming that f is Riemann-integrable (in the standard sense), this sequence $S = \langle S_j \rangle$ of Riemann-sums converges to the Riemann integral $\int_{\mathbb{U}} f(x) dx$.

Theorem 17 (Riemann Sum) *The Riemann-integral for a Riemann-integrable function on the unit interval is the standard part of the “hyper-average” of f , over a hyperfinite equi-partition $\mathcal{T}$ as above:*

$$\int_{\mathbb{U}} f(x) dx = \text{st} \left(\frac{1}{\mathcal{N}} \sum ({}^*f(t) : t \in \mathcal{T}) \right).$$

Remark 6 For the improper Riemann integral $\int_0^1 (dx)/\sqrt{1-x}$, the formula is valid.

Definition 34 Let now $f: \mathbb{R} \mapsto \mathbb{R}$ be a function, bounded, and compactly supported, which means $f(x) = 0$ for x outside a finite interval, then we may define (the “integral over $\mathbb{R}$ ”).

$$\begin{aligned} \int f(x) dx &:= \text{st} \left(\frac{1}{\mathcal{N}} \pi_{\mathbb{U}} \langle S_j \rangle \right); \\ \text{where } S_j &:= \sum (f(x) : x \in T_j); \\ T_j &= \left\{ \frac{k}{N_j} : k \in \mathbb{Z}, -N_j^2 \leq k < N_j^2 \right\}; \end{aligned}$$

And then for any bounded function f defined on $\mathbb{R}$:

$$\int_a^b f(x) dx := \int f(x) * 1_{[a \dots b]}(x) dx.$$

The definition is valid for any function f as long as it is bounded. (This is just like the Banach-limit for a bounded sequence. Axiom of choice is also involved in the latter!) The Riemann integrability is relevant only to prove that this “integral” is independent of the Riemann-summation scheme. We will now skip all these matter, mentioning only that, as is easily seen:

$$\begin{aligned} (a < b < c) &\Rightarrow \int_a^c f(x) dx \\ &= \int_a^b f(x) dx + \int_b^c f(x) dx. \end{aligned}$$

This justifies the usual definition that.

$$\int_a^b f(x) dx := - \int_b^a f(x) dx, \quad \text{for } b < a.$$

Theorem 18 (Fundamental Theorem of Calculus)

- (1) If f is continuous on an interval containing a point a , and $g(u) := \int_a^u f(x) dx$, then g is continuous, and for a continuity point b of f , the function g is differentiable at b , with $g'(b) = f(b)$.
- (2) If f is derivable on the interval $[a \dots b]$, so that the derivative f' is continuous, hence Riemann integrable, then.

$$\int_a^b f'(x) dx = f(b) - f(a).$$

NS-ODE

Theorem 19 (Cauchy–Peano) *If $f: (x, t) \in \mathbb{R} \times [0 \dots 1] \mapsto f(x, t) \in \mathbb{R}$ is bounded and continuous, then, for any “initial-position” $a \in \mathbb{R}$, there is a “solution-path”*

$$y : t \in [0 \dots 1] \mapsto y(t) \in \mathbb{R},$$

such that: it is continuous, and

$$y(0) = a, \quad y'(t) = f(y(t), t), \quad \forall t \in [0 \dots 1].$$

Proof The polygonal approximation is the following well-known scheme

$$y = \lim_{j \rightarrow \infty} y_j :$$

Take a natural number N_j , and divide the time-interval into N_j equal subintervals. In the duration of k th subinterval $(k-1)/N_j \leq t < k/N_j$, the velocity is approximated by the velocity-field at the latest position: $y_j(t) := f(y((k-1)/N_j), (k-1)/N_j)$, and consequently there is a recursion:

$$y_j\left(\frac{k}{N_j}\right) := y_j\left(\frac{k-1}{N_j}\right) + \frac{1}{N_j} * f\left(y_j\left(\frac{k-1}{N_j}\right), \frac{k-1}{N_j}\right); \quad k = 1, 2, \dots, N_j.$$

The standard way is to take a “limiting point” of the sequence of approximate solutions (which are polygonal = “piecewise-linear”), $y_j(t)$, $j = 1, 2, 3, \dots$. They are considered as points of the path-space (=function space, with uniform topology). Ascoli’s compactness criterion ensures the existence of such a limiting point y_∞ . From the uniform convergence

$$\lim_{j \rightarrow \infty} y_j(t) = y_\infty(t),$$

$$\lim_{j \rightarrow \infty} f(y_j(t), t) = f(y_\infty(t), t),$$

it follows easily the sought-for identity:

$$y_\infty(t) = a + \int_0^t f(y_\infty(s), s) ds.$$

The standard proof is translated into the NS proof thus: With $N_j \uparrow \infty$, we have an internal hyper-natural number (unlimited, of course), $\mathcal{N} = \pi_{\mathcal{U}}(N)$, $N := \langle N_j \rangle$, together with the internal set of partition-points on the NS time-interval:

$$\mathcal{T} := \left\{ \frac{k}{\mathcal{N}} : k \in \mathbb{Z}, \quad 0 \leq k \leq \mathcal{N} \right\}.$$

For “time” $\tau = \frac{k}{\mathcal{N}}$ on this hyperfinite set, define the $\mathcal{N}$ th approximate “position” $Y(\tau)$ to be

$$Y\left(\frac{k}{\mathcal{N}}\right) = a + \frac{1}{\mathcal{N}} * \sum_{i=0}^{k-1} f\left(Y\left(\frac{i}{\mathcal{N}}\right), \frac{i}{\mathcal{N}}\right).$$

Finally take

$$y(t) := \text{st}\left(Y\left(\frac{\text{floor}(t * \mathcal{N})}{\mathcal{N}}\right)\right).$$

The above NS summation-multiplied-by-infinitesimal goes to Riemann integral:

$$y(t) = a + \int_0^t f(y(s), s) ds.$$

□

It is well-known that under the very general condition of boundedness and continuity, the solution is not unique, reflecting the arbitrariness in the choice of one limiting point from a compact set. Of course this corresponds to the arbitrariness of the choice of the unlimited $\mathcal{N} \in {}^*\mathbb{N}$, or rather, the hyper-finite partition $\mathcal{T}$.

General Idea of NSA**The Set-Terminology**

Ever since G. Cantor invented the Mengenlehre, and after the Hilbertian axiomatization (r) evolution, the naive set-theory became the common language of mathematics. Everything is (to be expressed in terms of) set! As an example, the color “blue” may be defined as “the set of all blue things”. (I’d like to make this as my first teaching of pedagogy.) If this sentence is hardly

comprehensible, then an easier explanation is as follows. Suppose I say: “This stone ω is blue”. Then what I mean is the mathematical sentence “ $\omega \in B$ ”, where B is a conveniently collected set of blue stones. Of course, B should be a subset of a convenient “total set” Ω of stones. With this primitive idea of ‘membership relation’, simple set-notions fit in with our daily logic. Precisely: if a is the sentence “ $x \in A$ ”, and b is “ $x \in B$ ”, then the logical combinations of assertions correspond to the Boolean combinations of sets in the well-known way:

“ a or b ” is “ $x \in A \cup B$ ”;
 “ a and b ” is “ $x \in A \cap B$ ”;
 “not b ” is “ $x \in \Omega \setminus B$ ”.

Here a convenient total set Ω should be available for the logical analysis. From this thinking, a set $\mathcal{A}$ of subsets of Ω is called a Boolean algebra of sets, if it is closed under the three set-operations mentioned above: $\cup$, $\cap$, and “complementation” (set-difference by the total set). My second example of the “triumph of naive set theory” is the axiomatization of probability theory by A. N. Kolmogorov. Its publication was 10 years after N. Wiener’s construction of Wiener process. Wiener wanted to talk about the “probabilistic behavior of the path W of a particle”. By a path is meant a function from the time-axis to the state-space, $W(t)$ signifying “the position at time t ”. Here the time-axis is the half-line $[0 \cdots +\infty)$, and the state-space is the real line $\mathbb{R}$. And Wiener requires that the path is continuous, and starts from the origin: $X = 0$. In the view point of Kolmogorov, the phrase “probabilistic behavior of the path W of a particle” should be changed to “behavior of a random particle”. “A random particle” just means that it is “one chosen from many”; these many form a set. (This set, the so-called probability space, is always an imagined set. It is almost always an infinite set, so never “concrete” for students of probability theory!) The picture is: Lady Luck collects a big set Ω of all particles (=chance-lings). For each particle $\omega \in \Omega$, its path is $W(\cdot, \omega)$, mapping time t to its position

$W(t, \omega)$ at time t . (So Wiener process W is a “two-variable function”, its domain of definition is the set-product $\mathbb{R}_+ \times \Omega$. Textbooks of probability theory always suppress the variable ω which signifies the randomness.) The success of Kolmogorov’s axiomatization is simply this: it makes precise what kind of “questions about probability” one can ask. Now probability comes in, only because Lady Luck chooses one chance-ling ω from all conceivable. (Then this chosen one is the “reality”). We are interested (only) in the assertions of this form “ $\omega \in B$ ”. We can only ask questions of this kind: “What is the probability of the event $\omega \in B$?” And that only for some kind of set B . The answer $\mu(B)$ is the probability predestined by Her. For what kind of $B \subset \Omega$ is $\mu(B)$ meaningful is also Her designation. Then we simply call B an event. So $\mu(B)$ is now the probability of event B . (Instead of the “event $\omega \in B$ ”). Since we can do Boolean combinations of assertions, correspondingly, the set $\mathcal{A}$ of all events should be a Boolean set-algebra. But in probability theory, always the stronger condition of sigma-completeness is required: $\mathcal{A}$ is closed under countable union, thus called a σ -algebra.

Definition 35 ((Kolmogorov) Probability Space) A probability space is a nonempty set Ω , together with a measure structure $\langle \mathcal{A}, \mu \rangle$, where $\mathcal{A}$ is a σ -algebra on Ω , and $\mu: \mathcal{A} \rightarrow \mathbb{U}$ satisfies the countable-additivity condition: If $\forall n, B_n \in \mathcal{A}$ and $\forall m \neq n, B_m \cap B_n = \emptyset$, then

$$\mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n).$$

Returning to Wiener’s process. Lady Luck should allow assertions $\omega \in B$ of the form $W(t, \omega) \in (c_1 \dots c_2]$, where time $t \in [0 \cdots +\infty)$ and interval $(c_1 \dots c_2]$ are arbitrary. With this condition satisfied, the two-variable function $W: \langle t, \omega \rangle \mapsto W(t, \omega)$ is called a stochastic process on the probability space $(\Omega, \mathcal{A}, \mu)$. Actually the central questions asked by (Bachelier and) Wiener are events of the following forms:

$$B := \left\{ \omega \in \Omega : \text{ for } k = 1, 2, \dots, n, W(t_k, \omega) - W(s_k, \omega) \in A_k \right\},$$

where

$$s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n; \\ A_1, A_2, \dots, A_n, \text{ are intervals}$$

Theorem 20 (Wiener) *There exists a probability space $(\Omega, \mathcal{A}, \mu)$, and a stochastic process W on it, such that: for each $\omega \in \Omega$, the sample-path $W(\cdot, \omega)$ is a continuous function on $[0 \cdots + \infty)$; and, for set B as above,*

$$\mu(B) = \prod_{1 \leq k \leq n} \frac{1}{\sqrt{2\pi(t_k - s_k)}} \int_{A_k} e^{\frac{-x^2}{2(t_k - s_k)}} dx.$$

Naive Set Operations

As we have explained above, a mathematical system is a set with structure, which is a finite set of items of functions, relations, or functions. We now explain that starting from a ground set, there is a super-structure, which will contain all conceivable structures. Naive as we are, sometimes we just have to be very careful. So we give a very careful definition of couple and relation, which will be a foundation of all that follows.

Definition 36 (Couple and Cartesian Product)

For two non-empty sets A, B , the Cartesian product is defined as

$$A \times B := \{ \langle a, b \rangle : a \in A, \text{ and } b \in B \}.$$

Here “a couple”, or “an ordered pair”, is defined by

$$\langle a, b \rangle := \{ \{a\}, \{a, b\} \}.$$

The ordered n -tuple, for $n > 2$, may be defined recursively as

$$\langle x_1, x_2, \dots, x_n \rangle := \langle \langle x_1, x_2, \dots, x_{n-1} \rangle, x_n \rangle.$$

Note that we could write the “1-tuple” as $\langle a \rangle := a$. Then write:

$$A^n := \{ \langle x_1, x_2, \dots, x_n \rangle : x_1, x_2, \dots, x_n \in A \}.$$

Thus $A \times B \times C$ is defined to be $(A \times B) \times C$. (And so forth!)

The key to this definition is the observation that

$$(\langle x, y \rangle = \langle u, v \rangle) \text{ iff } ((x = u) \text{ and } (y = v)).$$

Here it is to be emphasized that when forming a set, multiple enumeration of any element just counts once, but does no harm. And in the case of forming the pair $\langle a, b \rangle$, when $a = b$, the pair is reduced to $\langle a, a \rangle := \{ \{a\} \}$, and this is a singleton set whose only element is the singleton set $\{a\}$.

Definition 37 (Relation) If $R \subset A \times B$, then R is a relation from A to B . And in case $A = B$, it is called a relation on A . The domain of R is the set

$$\text{Dom}(R) := \{ x \in A : \exists y \in B, \langle x, y \rangle \in R \}.$$

If $\langle x, y \rangle \in R$, sometimes we write (the “infix notation”!) xRy , still sometimes (the “prefix notation”!) $R(x, y)$. Suppose R and S are two relations, and $R \subset S$, then R is a restriction of S , and S an extension of R .

Definition 38 (Function and Strict-Evaluation)

If f is a relation from A to B , and, for any $x \in A$, there is exactly one $y \in B$, such that $\langle x, y \rangle \in f$, then we call f a function. For a function f and an element x in its domain of definition, the ‘evaluation’ gives the value $f(x)$ of x under f . So this may be generalized in the following way: if s, R are fixed, and there is exactly one t , such that: $\langle s, t \rangle \in R$, we then write:

$$t := R \uparrow^* s.$$

Definition 39 (n -ary Relation) If $R \subset A^n$, R is an n -ary relation on A . How to define a two-variable function? Let us take the arithmetic functions $+$ (addition), and $*$ (multiplication), as examples. They are from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$, are usually infix-written, thus

$x + y$ is precisely $\text{add} \ulcorner \langle x, y \rangle$;
 $x * y$ is precisely $\text{multiply} \ulcorner \langle x, y \rangle$.

Definition 40 (Power Set, Exponential Set) The set of all subsets of a set A , (including the empty set $\emptyset$, and the total set A), is denoted by $\mathbb{P}(A)$, and called the power-set of A . Starting with $\mathbb{P}^1(A) := \mathbb{P}(A)$, iteration of this power-set construction gives

$$\mathbb{P}^{n+1}(A) = \mathbb{P}(\mathbb{P}^n(A)), \quad \text{and we add } \mathbb{P}^0(A) = A.$$

Consider the set $A = \mathbb{D}_n := \{0, 1, \dots, n-1\}$, which has cardinality $n \in \mathbb{N}$, then $\mathbb{P}(A)$ has cardinality 2^n . We think “power-set” is an unfortunate terminology, in our opinion. A better one is the “exponential set”, and a better notation is 2^A . 2 is, to be sure, the natural base for the discrete mathematics. Or, if (2 being identified with the set $\mathbb{D}_2$ above), $f \in 2^A$ is interpreted as a function from A to $\mathbb{D}_2$, it is surely the indicator function of a subset of A . On the other hand, sets like $A^2, A^3, \dots$, should be called the power sets of A . (A to the second, third, power, $\dots$) It is generally known that in naive set theory, it is illegal to talk about “the set of all sets”. A convenient rule is that any set cannot be an element of itself: $A \notin A$. For any set A , then $A \neq \{A\}$, and indeed the latter is a singleton set. Now there is no entity as a set containing everything, there are sometimes some entity which can not be a set.

Definition 41 (Ur-element) Sometimes a non-empty set S may be declared as a ground-set, and its elements declared as ur-elements. A mathematical entity x is declared as an ur-element (or individual), if a sentence $y \in x$ is never allowed true. (A mathematical entity is either an ur-element or a set).

An ur-element is just like a sodium ion which is declared decomposable in high-school or freshman Chemistry. This is the most convenient concept for the students.

Definition 42 (Super-Structure) Suppose S is a set of ur-elements. We define $S_0 := S$, and then recursively:

$$S_n := S_{n-1} \cup \mathbb{P}(S_{n-1}).$$

The set

$$S_\infty := \bigcup_{j \geq 0} S_j$$

is called the super-structure over the ground-set S . Any $x \in S_\infty$ is assigned a “rank”

$$\min\{j : x \in S_j\}.$$

This rank is positive, iff x is a set.

Notation 2 The following are the symbols for logical constructions:

- The symbol $\neg$ is read as “the negation of”.
- The symbol $\wedge$ is read as “and”.
- The symbol $\vee$ is read as “or”.
- The symbol $\Rightarrow$ is read as “imply”.
- The symbol $\exists$ is read as “there exists”.
- The symbol $\forall$ is read as “for all”.

Theorem 21 *Every concept in mathematics can be explained in terms of set. In mathematical discussion, we take a suitable ground set S , build up the super-structure S_∞ , then the simplest kinds of assertions are: “ $a \in b$ ”, or “ $a = b$ ”, where a and b are elements of S_∞ . From these simple sentences more complicated sentences are formed with the aid of logical symbols.*

Example 7 (Group) Consider a quadruple $\langle G, e, \Lambda, * \rangle$, where G is a set, $e, \Lambda, *$ are respectively 0-, 1-, 2-variable functions with a triple $\langle e, \Lambda, * \rangle$, on G . The group axiom for this system is the sentence:

$$\begin{aligned} & (\forall x \in G)(\forall y \in G)(\forall z \in G)((x * (y * z) \\ & = ((x * y) * z) \wedge ((e * x) = x) \wedge (\widehat{x} * x = e)). \end{aligned}$$

This is too formal. But I hope by now you will have no real obstacle in understanding and accepting it. We see that if G is contained in the ground-set S_0 , then $e, \Lambda, *$ are all in the superstructure S_∞ with ranks 0, 3, 5, respectively. It is important to realize that group theory should be defined as the study of the true-or-false of those sentences involving only the group structure. Suppose $\Phi : G \mapsto H$ is group-homomorphism. Then a sentence (about group) true in G will be true in $\Phi[G] \subset H$, after the sentence is transferred (or translated, or reinterpreted), by Φ . Commutativity is an instance:

$$(\forall x \in G)(\forall y \in G), \quad (x * y = y * x).$$

It is possible that G is commutative, and H is not, but the homomorphic image $\Phi[G](\not\subseteq H)$ is commutative. Any sentence (about group) valid in G is valid in $\Phi(G)$.

Theorem 22 (NSA Idea) *The star-extension from the superstructure to the hyper-universe is a “homomorphism”.*

The key ideas of NSA, in the superstructure approach is the following:

- From a ground set S_0 of ur-elements, we can build up set-theoretically the super-structure

$$S_\infty := \bigcup_{j \geq 0} S_j,$$

the “standard universe”. All the mathematical entities arising from S_0 are in it.

- Associated with S_0 , there is the corresponding non-standard ground-set $*S_0 = S_0^{\mathfrak{U}}$ of ur-elements. But

$$S_\infty^{\mathfrak{U}} := \bigcup_{n \geq 0} S_n^{\mathfrak{U}},$$

the superstructure of $S_0^{\mathfrak{U}}$, is not interesting, because most of the entities in it are not very interesting: they are external to us. Instead of it, and inside of it, the hyper-universe (or non-standard universe)

$$*S_\infty := \bigcup_{n \geq 0} *S_n,$$

of internal entities, is of central importance in NSA. Entities in $S_\infty^{\mathfrak{U}} \setminus *S_\infty$ are then called external.

- There is an injective mapping from S_∞ into $*S_\infty$, called the star-extension. Every standard entity $A \in S_\infty$ is star-mapped to its star-extension $*A \in *S$, which is its (non-standard or) hyper version, or star version in the hyper-universe. All these are the (hyper versions of) standard entities. Other entities of the hyper-universe $*S_\infty$ are internal but non-standard.
- There is a Transfer Principle, which says that the star-embedding is the all-encompassing “injective homomorphism”: If a sentence is true in the standard interpretation in terms of the standard universe, then this sentence holds true in the non-standard interpretation in terms of the hyper-universe. By “non-standard interpretation” we mean that all the constant entities occurring in the sentence are to be replaced by their star-extensions. (So a quantifying clause “ $\forall x \in A$ ” is replaced by “ $\forall x \in *A$ ”).

The Ultra-Power Construction

Note 3 The notation S_0, S_n, S_∞ have the meaning as defined above. (So S_0 is a ground-set of ur-elements). There will be only another set $S_0^{\mathfrak{U}}$ of ur-elements! Any appearance like $Z_n, *S_n$, shall not be interpreted as part of superstructure construction! $\mathfrak{U}$ is a free ultra-filter on a set I of indices, to which all the phrases “ultimate” refer.

- Let Z_n be the set of “sequences” $f = \langle f_j \rangle$ such that for $j, f_j \in S_n$. The sequence Z_n is increasing with n , with union

$$Z_\infty = \bigcup_{n \geq 0} Z_n.$$

- For each $n \in \mathbb{N}_0$, on the set Z_n , we define an equivalence-relation $\sim^{\mathfrak{U}}$ (as before), by: $(f \sim^{\mathfrak{U}} g)$ iff:

$$f_j = g_j \quad \text{for ultimate } j.$$

Obviously we see that attaching a suffix n to the equivalence relation is hardly necessary: $\overset{\mathfrak{u}}{\sim}$ for Z_{n+1} is an extension of $\overset{\mathfrak{u}}{\sim}$ for Z_n .

- So we may define $S_0^{\mathfrak{u}} = \pi_{\mathfrak{u}}(Z_0)$ to be the quotient set of Z_0 under this equivalence relation $\overset{\mathfrak{u}}{\sim}$, i.e., the set of all equivalence-classes of Z_0 . Now we declare that $S_0^{\mathfrak{u}}$ be a ground-set of ur-elements, so that a formula like $y \in x$ is for $x \in S_0^{\mathfrak{u}}$ is illegal. We then build up the super-structure over $S_0^{\mathfrak{u}}$ as before:

$$S_{\infty}^{\mathfrak{u}} := \bigcup_{n \geq 0} S_n^{\mathfrak{u}}, \quad S_{n+1}^{\mathfrak{u}} := S_n^{\mathfrak{u}} \cup \mathbb{P}(S_n^{\mathfrak{u}}).$$

- The canonical projection from Z_0 into $S_0^{\mathfrak{u}}$ (surjectively) being here denoted by $\pi_{\mathfrak{u}}$, we now recursively extend it to a mapping

$$\pi_{\mathfrak{u}} : f \in Z_n \mapsto \pi_{\mathfrak{u}}(f) \in S_n^{\mathfrak{u}}, \quad \forall n \in \mathbb{N}.$$

After $\pi_{\mathfrak{u}}(g) \in S_{n-1}^{\mathfrak{u}}$ are defined for every $g \in Z_{n-1}$, we then define $\pi_{\mathfrak{u}}(f)$ for $f \in Z_n \setminus Z_{n-1}$ to be

$$\pi_{\mathfrak{u}}(f) := \left\{ \pi_{\mathfrak{u}}(g) : g \in Z_{n-1}, \text{ and } g_j \in f_j, \right. \\ \left. \text{for ultimate } j \right\}.$$

It is easy to see that $\pi_{\mathfrak{u}}$ is well-defined on $S_{\infty}^{\mathfrak{u}}$, and, by this definition, all over each $S_n^{\mathfrak{u}}$, and thus over $S_{\infty}^{\mathfrak{u}}$, we have

$$(\pi_{\mathfrak{u}}(f) = \pi_{\mathfrak{u}}(h)) \Leftrightarrow (f \overset{\mathfrak{u}}{\sim} h).$$

The proof of this (and similar ones below) is tedious but easy consequence of the dichotomy condition of the ultra-filter.

- Finally we have the hyper-universe of internal entities:

$${}^*S_{\infty} := \bigcup_{n \geq 0} {}^*S_n; \quad {}^*S_n := \{\pi_{\mathfrak{u}}(f) : f \in Z_n\} \subset S_n^{\mathfrak{u}}.$$

Any set $B \in S_{\infty}^{\mathfrak{u}} \setminus {}^*S_{\infty}$ is called strictly external.

- And we see that (just as what we have done before for $I = \mathbb{N}$):

$$\begin{aligned} \pi_{\mathfrak{u}}\langle A_j \rangle \cap \pi_{\mathfrak{u}}\langle B_j \rangle &= \pi_{\mathfrak{u}}\langle A_j \cap B_j \rangle; \\ \pi_{\mathfrak{u}}\langle A_j \rangle \cup \pi_{\mathfrak{u}}\langle B_j \rangle &= \pi_{\mathfrak{u}}\langle A_j \cup B_j \rangle; \\ \pi_{\mathfrak{u}}\langle A_j \rangle \setminus \pi_{\mathfrak{u}}\langle B_j \rangle &= \pi_{\mathfrak{u}}\langle A_j \setminus B_j \rangle; \end{aligned}$$

And the set of all internal sets form a Boolean set-ring.

- For $\{f, g, h\} \subset Z_{\infty}$,

$$\begin{aligned} \pi_{\mathfrak{u}}h &= \{\pi_{\mathfrak{u}}f, \pi_{\mathfrak{u}}g\} \quad \text{iff } h_j = \{f_j, g_j\}, \\ &\quad \text{for ultimate } j; \\ \pi_{\mathfrak{u}}h &= \langle \pi_{\mathfrak{u}}f, \pi_{\mathfrak{u}}g \rangle \quad \text{iff } h_j = \langle f_j, g_j \rangle, \\ &\quad \text{for ultimate } j; \\ \pi_{\mathfrak{u}}h &= \pi_{\mathfrak{u}}f \uparrow \pi_{\mathfrak{u}}g \quad \text{iff } h_j = f_j \uparrow g_j, \\ &\quad \text{for ultimate } j. \end{aligned}$$

- For any $n \in \mathbb{N}_0$, if $A \in S_n$, it may be identified with the “sequence” with constant term A , i.e., we may consider $S_n \subset Z_n$. Then $\pi_{\mathfrak{u}}(A) \in {}^*S_n$ is written as *A , and called a standard internal entity, or a star version of A , in the hyper-universe, as was mentioned before. Also note that: though $S_{\infty} \notin S_{\infty}$, and ${}^*S_{\infty} \notin {}^*S_{\infty}$, the star notation is still compatible.

Remark 7 If $b \in A \in S_{\infty}$, then the standard entity ${}^*b \in {}^*A$. That is

$${}^{\sigma}A := \{{}^*b : b \in A\} \subset {}^*A.$$

All standard members in *A are gathered into the set ${}^{\sigma}A$. Indeed when A is a finite set, ${}^*A = {}^{\sigma}A$. But when A is an infinite set, the right side above is properly contained in the left-side set. That is: the set *A must contain non-standard elements. Then both ${}^{\sigma}A$ and ${}^*A \setminus {}^{\sigma}A$ are external sets.

This is the source of all confusions!

- Internality is transitive in the sense that

$$(x \in y \text{ and } y \in {}^*S_{\infty}) \Rightarrow (x \in {}^*S_{\infty}).$$

- For entities A, B, C in S_{∞} ,

$$\begin{aligned}
A = B & \text{ iff } {}^*A = {}^*B; \\
A \in B & \text{ iff } {}^*A \in {}^*B; \\
C = \langle A, B \rangle & \text{ iff } {}^*C = \langle {}^*A, {}^*B \rangle; \\
C = A \uparrow B & \text{ iff } {}^*C = {}^*A \uparrow {}^*B;
\end{aligned}$$

For sets A, B , in $S_\infty \setminus S_0$,

$$\begin{aligned}
{}^*(A \cup B) &= {}^*A \cup {}^*B; \\
{}^*(A \cap B) &= {}^*A \cap {}^*B; \\
{}^*(A \setminus B) &= {}^*A \setminus {}^*B; \\
{}^*(A \times B) &= {}^*A \times {}^*B;
\end{aligned}$$

Theorem 23 (Transfer Principle) *If a sentence is true in the standard universe S_∞ , then its star-version is true in the hyper-universe ${}^*S_\infty$.*

All simple mathematical sentences are assertions about either the identity of two mathematical entities, or the membership of one against the other. By the use of “logical connectives”, more complicated assertions are then constructed. So the principle means exactly the whole set of “homomorphisms” displayed above.

Theorem 24 (Standardization Principle) *Let $A \in S_\infty$ be a standard entity, and ${}^*A \in {}^*S_\infty$ its star-version. Let $B \in S_\infty^{\text{st}} \setminus S_0^{\text{st}}$ be any non-standard set. Then there is $C \in S_\infty$, such that: for any $x \in S_\infty$,*

$$({}^*x \in {}^*C) \text{ iff } ({}^*x \in ({}^*A \cap B)).$$

C is called the standard part of A . The proof is simple: C is the union of all subsets $D \subset A$, such that

$$({}^*x \in {}^*D) \Rightarrow ({}^*x \in ({}^*A \cap B)).$$

Definition 43 (Concurrency Relation) A relation $R \in S_\infty \setminus S_1$ is called concurrent, if: for any finite number of elements $a_1, a_2, \dots, a_n$ in the domain of R , there is one $b \in S_\infty$, such that

$$\langle a_i, b \rangle \in R, \quad \text{for } i = 1, 2, \dots, n.$$

Example 8 An obvious example of concurrency relation is the “less than” relation $<$ in the set $\mathbb{N}$. In

this case, there is an unlimited hyperinteger ω bigger than all standard integer $a \in \mathbb{N}$, that is: $\forall a \in \mathbb{N}, a < \omega$. All we need, in this particular example, is that the indicial set I has infinite cardinal.

More generally, we need an index set I with big enough cardinal. Then the following Concurrency Theorem is valid. (But the proof is bigger than our concern).

Theorem 25 (Idealization Principle) *Let $R \in S_\infty$ be a concurrent relation. Then there is $b \in {}^*S_\infty$, such that for all $a \in \text{Dom}(R)$, $\langle a, b \rangle \in R$.*

Note 4 (Internal Set Theory) There are now some axiomatic approach of NSA. In the internal set theory proposed by E. Nelson, the notion of “standardness” is considered as fundamental and undefined, and the three principles of idealization, standardization, and transfer, are taken as axioms, (in addition to the Fraenkel–Zermelo axioms with Axiom of Choice), for the NSA. This is very elegant, though not our choice in this article.

Some Consequences of Transference

Example 9 (Well-Ordering in ${}^*\mathbb{N}$) The axiom of mathematical induction for $\mathbb{N}$ is equivalent to the well-ordering property of $\mathbb{N}$:

$$\begin{aligned}
((\forall A \subset \mathbb{N}) \wedge (A \neq \emptyset)) &\Rightarrow (\exists m \in A, \\
&(\forall n \in A, m \leq n)).
\end{aligned}$$

“Every non-empty subset A of $\mathbb{N}$ has a minimum element”.

The transferred sentence is not:

$$((\forall A \subset {}^*\mathbb{N}) \wedge (A \neq \emptyset)) \Rightarrow (\exists m \in A, (\forall n \in A, m \leq n)).$$

“Every non-empty subset A of ${}^*\mathbb{N}$ has a minimum element”. Instead, the correct “translation” of this sentence is “Every non-empty internal subset A of ${}^*\mathbb{N}$ has a minimum element”. Why? We may write the original premise as $\forall A \in \mathbb{P}(\mathbb{N})$, then, in the transference, not only the set $\mathbb{N}$ is star-extended to ${}^*\mathbb{N}$, but also its power-set is star-extended to the standard set (in non-standard version), ${}^*\mathbb{P}(\mathbb{N})$.

$$((\forall A^* \subset {}^*\mathbb{P}(\mathbb{N})) \wedge (A \neq \emptyset)) \\ \Rightarrow (\exists m \in A, (\forall n \in A, m \leq n)).$$

Another way of thinking is to star-transform the original premise “ $\forall A \subset \mathbb{N}$ ” into “ $\forall A^* \subset {}^*\mathbb{N}$ ”. (That is: the subset-relation gets star-extended.) In the ultra-power representation used before, $A = \pi_{\mathcal{U}} \langle A_j \rangle$, then take $m_j = \min(A_j) \in \mathbb{N}$, which exists because of the well-ordering of $\mathbb{N}$, and $\mathbb{N} \supset A_j \neq \emptyset$. Then $m := \pi_{\mathcal{U}} \langle m_j \rangle \in {}^*\mathbb{N}$ is the minimum of A sought for. The wrong transference gives us the well-ordering property of ${}^*\mathbb{N}$, which is obviously wrong, because the set of all unlimited hyper natural numbers ${}^*\mathbb{N} \setminus \mathbb{N}$ would have a minimum element ζ , which is absurd because $\zeta - 1$ would have been still an element of this set.

Corollary 2 *The set $\mathbb{N}$ is an external subset of ${}^*\mathbb{N}$. So is the complement ${}^*\mathbb{N} \setminus \mathbb{N}$.*

Example 10 (Order-Completeness of ${}^*\mathbb{R}$) The order-completeness of $\mathbb{R}$ says that: “Every non-empty subset A of $\mathbb{R}$ which is bounded above has a supremum (= sup = least upper bound)”.

$$((\forall A \subset \mathbb{R}) \wedge (A \neq \emptyset) \wedge (\exists b \in \mathbb{R}, (\forall x \in A, x \leq b)) \\ \Rightarrow ((\exists m \in \mathbb{R}, (\forall x \in A, x \leq b)) \\ \wedge (\forall r \in \mathbb{R}, ((\forall x \in A, x \leq r) \Rightarrow m \leq r))).$$

The transferred sentence is not the order-completeness of the field ${}^*\mathbb{R}$: “Every non-empty subset A of ${}^*\mathbb{R}$ which is bounded above has a supremum”. A ready counterexample is the set of infinitesimals $\mathbb{I}$. $\mathbb{I}$ is certainly bounded above by any strictly positive standard real number. But it does not have a “supremum ζ ”. Otherwise, ζ could only be positive. If it is a positive real number, then $\frac{\zeta}{2}$ is a smaller upper-bound. If it is a positive infinitesimal, then an infinitesimal, namely $2*\zeta$ could not be bounded by ζ . (We may also appeal to the theorem that there is only one order -complete field, up to a canonical isomorphism.) Again the correct transference should replace “subset” by “internal subset”: “In ${}^*\mathbb{R}$, every non-empty internal subset A which is bounded above has a sup (= least upper bound)”.

Corollary 3 (Overflow and Underflow) *Let A be an internal subset of ${}^*\mathbb{R}$.*

- (1): *If for any $n \in \mathbb{N}$, there is a limited $x \in A$, such that $x > n$, then there is unlimited positive $\eta \in A$.*
- (2): *If for any unlimited positive $\zeta \in A$, there is another unlimited $\xi \in A$, such that $\xi < \zeta$, then there is a limited positive $\eta \in A$.*
- (3): *Both subsets $\mathbb{R}(\subset {}^*\mathbb{R})$, and ${}^*\mathbb{R} \setminus \mathbb{R}$, of the standard set ${}^*\mathbb{R}$ (hyper-version) are external.*

Proof

- (1) We may assume that A is bounded above. But then with $\xi = \sup(A) \in {}^*\mathbb{R}$, ξ is unlimited. Therefore by the definition of lub, there exists an $\eta \in A$, with $\frac{\xi}{2} \leq \eta \leq \xi$.
- (2) Let z be the infimum (= inf = greatset lower bound) of the subset of positive elements of A , then z is limited. And by the definition of inf, there is $\eta \in A$, such that $z \leq \eta \leq z + 1$. $\square$

Theorem 26 (Saturation Principle) *Let $\langle A_n : n \in \mathbb{N} \rangle$ be a sequence of internal sets with the “finitely-intersecting property”:*

$$\forall n \in \mathbb{N}, \quad \bigcap_{1 \leq m \leq n} A_m \neq \emptyset.$$

Then the sequence is itself intersecting:

$$\bigcap_{m \in \mathbb{N}} A_m \neq \emptyset.$$

(For the proof, the countable indicial set $I = \mathbb{N}$ is enough.) Now for each $n \in \mathbb{N}$, $A_n = \pi_{\mathcal{U}} \langle A_{n,j} : j \in I \rangle$. And for this fixed $n \in \mathbb{N}$,

$$\bigcap_{m \in \mathbb{N}} A_m = \pi_{\mathcal{U}} \langle \bigcap_{1 \leq m \leq n} A_{m,j} : j \in I \rangle \neq \emptyset \in {}^*S_{\infty}.$$

Therefore for each $n \in \mathbb{N}$, and for ultimate j , $\bigcap_{1 \leq m \leq n} A_{m,j} \neq \emptyset$. (In particular, we may and will just assume: for all j , $A_{1,j} \neq \emptyset$.) For each $j \in I$, we define

$$k_j := \max\{n \in \mathbb{N} : \cap_{1 \leq m \leq n} A_{m,j} \neq \emptyset, \text{ and } n \leq j\};$$

Now for each $j \in I$, choose an $x_j \in \cap_{1 \leq m \leq k_j} A_{m,j}$. We see that $\xi = \pi_{\mathcal{U}} \langle x_j \rangle \in A_n$, for any $n \in \mathbb{N}$. Indeed for any $n \in \mathbb{N}$,

$$\begin{aligned} \{j \in I : x_j \in A_{n,j}\} &\supset \{j : k_j \geq n\} \\ &= \{j \in I : j \geq n\} \cap \{j \in I : \cap_{m \leq n} A_{m,j} \neq \emptyset\} \in \mathcal{U}. \end{aligned}$$

Corollary 4 For a sequence $\langle A_n : n \in \mathbb{N} \rangle$ of internal sets, their (countably infinite) union is internal $\cup_{n \in \mathbb{N}} A_n \in {}^*S_\infty$, iff the union equals to a finite sub-union. I. e., iff there exists $m \in \mathbb{N}$, such that

$$\cup_{n \in \mathbb{N}} A_n = \cup_{1 \leq n \leq m} A_n.$$

Proof If $B := \cup_{n \in \mathbb{N}} A_n \in {}^*S_\infty$ (is internal), then: all $B_n := B \setminus A_n$ are internal, and $\cap_{n \in \mathbb{N}} B_n = \emptyset$, by de Morgan's law. Then the saturation principle requires that for a certain $m \in \mathbb{N}$, $\cap_{1 \leq n \leq m} B_n = \emptyset$. Or, $\cap_{1 \leq n \leq m} (B \setminus A_n) = \emptyset$, and $\cup_{1 \leq n \leq m} A_n = B = \cup_{n \in \mathbb{N}} A_n$. $\square$

Loeb Construction

In this section we provide the Loeb version of constructing a non-standard version of stochastic analysis (cf. (Loeb und Wolff 2000)).

Measure

Let a basic non-empty set Ω be fixed. We recall that a subset $\mathcal{A} \subset \mathbb{P}(\Omega)$ is called a Boolean algebra of sets, iff the following requirements are satisfied:

$$\begin{aligned} \Omega &\in \mathcal{A}; \\ (\forall x \in \mathcal{A})(\forall y \in \mathcal{A}), \quad &(x \cup y \in \mathcal{A}); \\ (\forall x \in \mathcal{A})(\forall y \in \mathcal{A}), \quad &(x \cap y \in \mathcal{A}); \\ (\forall x \in \mathcal{A})(\forall y \in \mathcal{A}), \quad &(x \setminus y \in \mathcal{A}). \end{aligned}$$

Or the last three sentences can be combined into one:

$$\begin{aligned} &(\forall x \in \mathcal{A})(\forall y \in \mathcal{A}), \\ &((x \cup y \in \mathcal{A}) \wedge (x \cap y \in \mathcal{A}) \wedge (x \setminus y \in \mathcal{A})). \end{aligned}$$

If $\mathcal{A}$ is closed under countable operations, it is then called a σ -algebra on Ω :

$$\begin{aligned} &\forall (x_n : n \in \mathbb{N}), \\ &((\forall n \in \mathbb{N}, x_n \in \mathcal{A}) \Rightarrow (\cup_{n \in \mathbb{N}} x_n \in \mathcal{A})). \end{aligned}$$

It is well-known that for any set $\mathcal{B}$ of subsets of Ω , there is uniquely a smallest σ -algebra $\mathcal{A} \supset \mathcal{B}$. This is the σ -algebra generated by $\mathcal{B} \in \mathbb{P}(\mathbb{P}(\Omega))$.

Definition 44 (Probability Content and Measure) Let $\mathcal{A}$ be a Boolean algebra of sets on a total set Ω . A function

$$\mu : A \in \mathcal{A} \mapsto \mu(A) \in \mathbb{R}$$

is called a content on $(\Omega, \mathcal{A})$, if the following conditions are satisfied:

$$\begin{aligned} (1): \quad &\forall x \in \mathcal{A}, \quad \mu(x) \geq 0; \\ (2): \quad &\forall x \in \mathcal{A}, \quad \forall y \in \mathcal{A}, (x \cap y = \emptyset) \\ &\Rightarrow (\mu(x \cup y) = \mu(x) + \mu(y)). \end{aligned}$$

If the following condition of “the continuity from above at the empty set” is satisfied, then μ is called a measure:

$$\begin{aligned} &\forall (x_n : n \in \mathbb{N}), ((\forall n \in \mathbb{N} (x_n \in \mathcal{A}) \wedge (x_{n+1} \subset x_n)) \\ &\wedge (\cap_n x_n = \emptyset)) \Rightarrow (\lim_{n \rightarrow \infty} \mu(x_n) = 0). \end{aligned}$$

An equivalent condition is: μ is σ -additive, in the sense that: if $\langle A_n \rangle$ is a disjoint sequence in $\mathcal{A}$, with union $\cup_{n \in \mathbb{N}} A_n \in \mathcal{A}$, then

$$\mu(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

If content (or measure) μ is normalized: $\mu(\Omega) = 1$, then it is called a probability content (or measure, respectively).

Theorem 27 (Extension of a Measure) If μ is a measure on $(\Omega, \mathcal{A})$, and $\mathcal{B}$ is the σ -algebra generated by $\mathcal{A}$, then there is a unique extension of μ to

a measure $\tilde{\mu}$ on $(\Omega, \mathcal{B})$. Further, let $\overline{\mathcal{A}}$ be the collection of sets of the form $A = (B \setminus D) \cup (D \setminus B)$, where $B \in \mathcal{B}$, and $D \subset C$, with $C \in \mathcal{B}$, $\tilde{\mu}(C) = 0$, and define then $\bar{\mu}(A) = \tilde{\mu}(B)$, then the definition is valid, and $\bar{\mu}$ is a measure on $(\Omega, \overline{\mathcal{A}})$. This extension of μ is the “completion of μ ”. It is “complete” in the sense that whenever $A \in \overline{\mathcal{A}}$, $F \subset A$, and $\bar{\mu}(A) = 0$, then $F \in \overline{\mathcal{A}}$ with $\bar{\mu}(F) = 0$.

Remark 8 (σ -Finiteness) The definition of a content or measure, is generally extended to allow $+\infty$ in the range. Then in the discussion the condition of σ -finiteness is is generally sufficient and necessary: there is a sequence $A_n \in \mathcal{A}$, with $\bigcup_{n \in \mathbb{N}} A_n = \Omega$, and $\mu(A_n)$ finite, for all n .

Loeb Measure

Definition 45 (Loeb Content) Let $\Omega \in {}^*S_\infty$ be an internal set, $\mathcal{A}$ be an internal set of subsets of Ω , which is closed under complementation and finite union. A Loeb content on $(\Omega, \mathcal{A})$ is an internal function $\mu : \mathcal{A} \Rightarrow {}^*\mathbb{R}$, which is a content on $\mathcal{A}$, i.e., it is finitely-additive and non-negatively valued. (But it is valued in ${}^*\mathbb{R}$, not real-valued.)

Theorem 28 (Loeb Measure) For a Loeb content space $(\Omega, \mathcal{A}, \mu)$, define the standard-part of μ as

$${}^\circ\mu : A \in \mathcal{A} \mapsto {}^\circ\mu(A) := {}^\circ(\mu(A)).$$

It is then s -additive, and its completion is the Loeb measure $L(\mu)$ of μ , defined on the Loeb algebra $L(\mathcal{A})$ consisting of those $B \subset \Omega$ which are “ μ -approximable”: For any positive real?, there are $A \in \mathcal{A}$, $C \in \mathcal{A}$, such that:

$$A \subset B \subset C, \quad \mu(C) - \mu(A) < \epsilon.$$

Indeed there is $D \in \mathcal{A}$, (called Loeb’s modification of B), such that

$$L(\mu)(B \Delta D) = 0.$$

Here Δ is the symmetric-difference:

$$B \Delta D := (B \setminus D) \cup (D \setminus B).$$

We omit the proof, but by the saturation property of internal sets, the continuity or s -additivity of μ is trivial. It should be emphasized that $L(\mu)$ is a real-valued measure; therefore the integrands of its integration are real-valued, but defined on Ω , hence is not internal.

Definition 46 (Measurable Functions) If temporarily we denote by $\mathcal{G}$ the set of all open sets of $\mathbb{R}$, (also called the topology of $\mathbb{R}$), and call elements of ${}^*\mathcal{G}$ star-open, then an internal function $F : \Omega \Rightarrow {}^*\mathbb{R}$ is called $\mathcal{A}$ -measurable, iff:

$$(G \in {}^*\mathcal{G}) \Rightarrow (F^{-1}(G) \in \mathcal{A}).$$

If $F := \pi_{\mathcal{U}} \langle F_j \rangle$, $\mu = \pi_{\mathcal{U}} \langle \mu_j \rangle$, we define naturally:

$${}^*\int F d\mu := \pi_{\mathcal{U}} \left\langle \int F_j d\mu_j \right\rangle.$$

This is valued in ${}^*\mathbb{R}$.

Remark 9 (Loeb Integral) If F is real-bounded, then

$${}^\circ \left({}^*\int F d\mu \right) = \int {}^\circ F dL(\mu),$$

which is then called the Loeb integral. The equality here is guaranteed by the S -integrability of the internal function F :

- (1): ${}^*\int |F| d\mu$ is limited.
- (2): If $\mu(A) = 0$, then ${}^*\int |F| d\mu = 0$.
- (3): If $F(\omega) = 0$ for all $\omega \in A$, then ${}^*\int |F| d\mu \approx 0$.

Now we turn to the integration of a real-valued function f on Ω , with respect to the real-valued measure $L(\mu)$.

Theorem 29 (Lifting of Function) Any $L(\mathcal{A})$ -measurable function f has a “lifting” $F : \Omega \mapsto {}^*\mathbb{R}$,

in the sense that: F is internal, $\mathcal{A}$ -measurable function, satisfying

$$L(\mu)\{\omega : {}^\circ F(\omega) \neq f(\omega)\} = 0.$$

Definition 47 (Normed Counting Measure) If $\Omega \in {}^*S_\infty$ is a hyperfinite set, and $\mathcal{A}$ the set of all internal set, the internal probability content μ ,

$$\mu(A) := \frac{{}^*\text{card}(A)}{{}^*\text{card}(\Omega)},$$

is called the normed counting measure on Ω .

Example 11 (Hyperfinite Equi-partition of the Unit-Interval) Let $\mathcal{N} \in {}^*\mathbb{N} \setminus \mathbb{N}$ be an unlimited hyperinteger, $\mathcal{T}$ be the equi-partition hyperset of the unit interval into $\mathcal{N}$ parts, and μ be the normalized uniform counting measure. Then we get the Loeb measure $L(\mu)$ and the Loeb algebra $L(\mathcal{A})$ on $\mathcal{T} \subset {}^*[0 \dots 1]$. The Loeb measure space $(\mathcal{T}, L(\mathcal{A}), L(\mu))$ is simply the pull-back of the standard Lebesgue measure space $(\mathbb{U}, \mathfrak{B}(\mathbb{U}), \lambda)$ by the standard-part map.

Here we have a dual of push-down and pull-back (or lifting). The push-down is via standard-part map $\text{st} : \mathcal{T} \mapsto \mathbb{U}$, which is a surjection, while its inverse st^{-1} furnishes the pull-back. The push-down of $L(\mathcal{A})$, by st , is by definition,

$$\{B \subset \mathbb{U} : \text{st}^{-1}[B] \in L(\mathcal{A})\}.$$

This is precisely the Lebesgue algebra $\overline{\mathfrak{B}(\mathbb{U})}$. And for a member B in this latter algebra, then the push-down of $L(\mu)$ will assign the measure value $L(\mu)(\{x \in \mathcal{T} : {}^*x \in B\})$. But we have already shown that when B is an interval, this value is simply its length $\lambda(B)$. This is enough to identify the pushed-down measure. Actually, the same kind of Loeb construction may be applied to the whole real axis $\mathbb{R}$, or the real half-axis $\mathbb{R}_{+}$, instead of $\mathbb{U}$ (or the analogous finite intervals). A little bit of attention should be paid to the s-finiteness, of course.

Anderson's Wiener Process

Definition 48 (Hyperfinite Time-axis) When $\mathcal{N} \in {}^*\mathbb{N} \setminus \mathbb{N}$ is fixed, $\mathcal{N} = \pi_{\mathcal{U}}\langle N_j \rangle$, we have the hyperfinite time-axis

$$\mathcal{T} := \pi_{\mathcal{U}}\langle T_j \rangle; T_j := \left\{ \frac{k}{N_j} : k \in \mathbb{Z}, 0 \leq k < N_j^2 \right\}.$$

As before, T_j is the interval $[0 \dots N_j]$ equi-partitioned into N_j^2 parts, each with length $1/N_j$. The right-end N_j is deliberately deleted, so that $\text{card}(T_j) = N_j^2$, while the augmented set is $T'_j := T_j \cup \{N_j\}$. The set T_j will be called rank k (discretized) time-axis. The left- and the right-approximations from the semi-axis $[0 \dots +\infty)$ to $\mathcal{T}$ is well-defined:

$$\begin{cases} \text{floor}_{\mathcal{T}}(c) = \pi_{\mathcal{U}}\left\langle \frac{\text{floor}(c * N_j)}{N_j} : j \in I \right\rangle; \\ \text{ceil}_{\mathcal{T}}(c) = \pi_{\mathcal{U}}\left\langle \frac{\text{ceil}(c * N_j)}{N_j} : j \in I \right\rangle. \end{cases}$$

Definition 49 (Hyperfinite Coin-Tossing) With $\mathcal{N} \in {}^*\mathbb{N} \setminus \mathbb{N}$, and $\mathcal{T} = \left\{ \frac{k}{\mathcal{N}} : k = 0, 1, 2, \dots, \mathcal{N} - 1 \right\}$ as above, we define:

$$\begin{aligned} \text{Coin} &:= \{-1, +1\}; \\ \Omega_j &:= \text{Coin}^{T_j}; \\ \Omega &:= \pi_{\mathcal{U}}\langle \Omega_j \rangle; \end{aligned}$$

Of course, $+1$ or -1 means “head” or “tail” of a coin toss. Any element $\omega_j \in \Omega_j$ is called a coin-particle of rank j , and an element $\omega \in \Omega$ is called a coin-particle (of hyper-finite rank). The normalized counting measure P_j on Ω_j is simple indeed, as is the one occurring most often in our high-school textbooks. (We will denote by $\mathbb{E}_{P_j}$ the expectation with respect to P_j). The normalized counting measure P is defined on the algebra of all internal subsets of the hyperfinite internal set Ω . Therefore the Loeb construction give us the Loeb measure $L(P)$, called the hyperfinite coin-

measure, on the Loeb algebra $L(\mathcal{A})$, which is called the hyper-finite coin algebra.

Definition 50 (The (Rank j) Random Walker)

For fixed j , and any rank j coin-particle $\omega_j \in \text{Coin}^{T_j}$, its path is the mapping $t \mapsto B_j(\omega_j, t) \in \mathbb{R}$, defined as follows. The domain of definition is

$$T'_j := \left\{ \frac{k}{N_j} : k = 0, 1, 2, \dots, N_j^2 \right\}.$$

And when $t = \frac{k}{N_j}$,

$$B_j(\omega_j, t) := \frac{1}{\sqrt{N_j}} \sum \left\{ \omega_j \left(\frac{m}{N_j} \right) : 0 \leq m < k \right\}.$$

By definition, the walk starts from the origin $B_j(\omega_j, 0) := 0$. Then recursively, at each time $t = m/N_j$, when the (j -) particle is located at the position $B_j(\omega_j, t)$, the coin-tossing $\omega_j(t)$ will decide the plus or minus direction of the next step, occurring in the time-interval, from t to $t + \Delta_j t$. Here the rank j

$$\text{time} - \text{unit is } \Delta_j t := \frac{1}{N_j};$$

$$\text{space} - \text{unit is } \Delta_j x := \frac{1}{\sqrt{N_j}}.$$

As a coin-particle $\omega_j \in \text{Coin}^{T_j}$ is randomly chosen, we may call it a random walker.

Definition 51 (Hyperfinite Random Walk)

Define entities in the hyper-universe:

$$B := \pi_{\mathcal{U}} \langle B_j \rangle;$$

$$B(\omega, t) := \sum \left\{ \frac{\omega(s)}{\sqrt{\mathcal{N}}} : s < t \right\};$$

$B: \langle \omega, t \rangle \in \Omega \times \mathcal{T} \mapsto B(\omega, t) \in {}^*\mathbb{R}$ is called the hyperfinite random walk.

Lemma 8 (Independent Increment (Finite-Rank)) For fixed j , if there are $2n$ elements of T'_j , arranged according to order:

$$s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n.$$

Then these n increments:

$$B_j(\cdot, t_1) - B_j(\cdot, s_1), B_j(\cdot, t_2) - B_j(\cdot, s_2), \dots, B_j(\cdot, t_n) - B_j(\cdot, s_n),$$

are independent, in the sense that: for any choice of intervals (or Borel sets) $A_1, A_2, \dots, A_n$,

$$P_j \left\{ \omega_j : B_j(\omega_j, t_k) - B_j(\omega_j, s_k) \in A_k, \forall k = 1, 2, \dots, n, \right\} \\ = \prod_{1 \leq k \leq n} P_j \left\{ \omega_j : B_j(\omega_j, t_k) - B_j(\omega_j, s_k) \in A_k \right\}.$$

Proof Coin-tossings are independent, by assumption. By transference,

Lemma 9 (Independent Increment (Hyperfinite-Rank)) If $2n$ elements of $\mathcal{T}$ are chosen and arranged according to order:

$$s_1 < t_1 \leq s_2 < t_2 \leq \dots \leq s_n < t_n.$$

Then the increments:

$$B(\cdot, t_1) - B(\cdot, s_1), \dots, B(\cdot, t_n) - B(\cdot, s_n),$$

are $\star$ -independent, in the sense that: for any choice of internal sets $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$,

$$P \left\{ \omega : B(\omega, t_k) - B(\omega, s_k) \in \tilde{A}_k, \text{ when } k = 1, 2, \dots, n \right\} \\ = \prod_{1 \leq k \leq n} P \left\{ \omega : B(\omega, t_k) - B(\omega, s_k) \in \tilde{A}_k \right\}.$$

Lemma 10 (Fourier and Moments Characteristic, Finite-Rank) For $t \in T_j$, and $y \in \mathbb{R}$,

$$\begin{aligned}\mathfrak{E}_{P_j}\left(e^{\sqrt{-1}\theta*B_j(t)}\right) &= \left(\cos\left(\frac{y}{\sqrt{N_j}}\right)\right)^{\frac{t}{N_j}}; \\ \mathfrak{E}_{P_j}(B_j(t)) &= 0; \\ \mathfrak{E}_{P_j}(B_j(t))^2 &= t; \\ \mathfrak{E}_{P_j}(B_j(t))^4 &= 3t^2 - 2\frac{t}{N_j}.\end{aligned}$$

Proof To calculate $\mathfrak{E}_{P_j}\left(e^{\sqrt{-1}\theta*B_j(t)}\right)$, by independence, we need only calculate the case of only one-step, i.e., for $t = \frac{1}{N_j}$. Then it is

$$\frac{1}{2}\left(e^{\sqrt{-1}\frac{y}{\sqrt{N_j}}} + e^{\sqrt{-1}\frac{-y}{\sqrt{N_j}}}\right) = \cos\left(\frac{y}{\sqrt{N_j}}\right).$$

The odd moments are zero for the even-symmetry of the distribution. Then we do, for the second moment,

$$(u_1 + u_2 + u_3 + \dots)^2 = \sum u_k^2 + 2\sum_{k<l} u_k u_l.$$

Also: for the fourth moment,

$$\begin{aligned}(u_1 + u_2 + \dots)^4 &= \sum u_k^4 + 6\sum_{k<l} u_k^2 u_l^2 \\ &\quad + \text{odd terms.}\end{aligned}$$

By transference, then: $\square$

Lemma 11 (Fourier and Moments Characteristic, Hyperfinite-Rank) For any $t \in \mathcal{T}$, and for any $y \in \mathbb{R}$,

$$\begin{aligned}\mathfrak{E}_P\left(e^{\sqrt{-1}y*B(t)}\right) &= \left(\cos\left(\frac{y}{\sqrt{\mathcal{N}}}\right)\right)^{\frac{t}{\mathcal{N}}}; \\ \mathfrak{E}_P(B(t)) &= 0; \\ \mathfrak{E}_P(B(t))^2 &= t; \\ \mathfrak{E}_P(B(t))^4 &= 3t^2 - 2\frac{t}{\mathcal{N}};\end{aligned}$$

Theorem 30 (Anderson) *The function b defined below is a Wiener process.*

$$\begin{aligned}b(\omega, t) &:= \text{st}(B(\omega, \text{ceil}_{\mathcal{T}}(t))); \\ b : \Omega \times [0 \dots \infty) &\mapsto \mathbb{R}.\end{aligned}$$

- As to the independence of the increments:

$$b(\cdot, t_1) - b(\cdot, s_1), \quad b(\cdot, t_2) - b(\cdot, s_2), \dots, b(\cdot, t_n) - b(\cdot, s_n),$$

this utilizes the Loeb's modification, and the standard-part transformation between b and B .

- As to the Gaussian nature of the increment $b(\cdot, t) - b(\cdot, s)$, which is distributed just like $b(\cdot, t - s)$, calculation of its Fourier characteristic is the most convenient way of proof. Now we have (by calculus, L'Hospital or not):

$$\lim_{n \rightarrow \infty} \cos^{\text{floor}(nt)}\left(\frac{y}{\sqrt{n}}\right) = e^{-\frac{y^2}{2}t}.$$

Therefore

$$\int e^{\sqrt{-1}yb(t)} dL(P) = e^{-\frac{y^2}{2}t}.$$

- As to the continuity of sample-paths, the usual method is to calculate the moments of increment. The discussion we will omit here.

A Future for Non-standard Analysis?

The ultrapower construction above is the semantic approach. Another approach is called the *syntactic approach* to non-standard analysis requires much less logic and model theory to understand and use. This approach was developed in the mid-1970s by the mathematician Edward Nelson (Nelson 1977, 1989), using a theory called the *internal set theory* which has a unary operation “standard” that makes non-standard analysis possible. Despite the elegance exhibited in non-standard analysis, some claim that it is not very useful and can only do reinterpretations or reproofs of previous results

like (Kamae 1982; van den Dries und Wilkie 1984). In the study of random walks and such, NSA still find some applications, such as (Albeverio et al. 1986).

Bibliography

- Albeverio S, Fenstad JE, Hoegh-Krohn R, Lindstrøm T (1986) Nonstandard methods in stochastic analysis and mathematical physics. *Bull Am Math Soc* 17(2): 385–389
- Bernstein A, Robinson A (eds) (1966) Solution of an invariant subspace problem of KT Smith and PR Halmos. *Pac J Math* 16(3):421–431
- Halmos P (1966) Invariant subspaces for polynomially compact operators. *Pac J Math* 16(3):433–437
- Kamae T (1982) A simple proof of the ergodic theorem using nonstandard analysis. *Israel J Math* 42(4): 284–290
- Keisler HJ (1986a) An infinitesimal approach to stochastic analysis. *J Symb Logic* 51(3):822–824. now included In: (1984) *Mem Am Math Soc* 297
- Keisler HJ (1986b) *Elementary calculus: an approach using infinitesimals*, 2nd edn. Now available at <http://www.math.wisc.edu/%7Ekeisler/calc.html>
- Kelley JL (1975) *General topology*. Springer, New York
- Loeb PA, Wolff MPH (2000) Nonstandard analysis for the working mathematician, *Math. Appl.*, vol. 510. Kluwer, Dordrecht
- Nelson E (1977) Internal set theory a new approach to nonstandard analysis. *Bull Am Math Soc* 83(6): 1165–1198; Also see <http://www.math.princeton.edu/%7Eenelson/books/1.pdf>
- Nelson E (1989) Radically elementary probability theory. *Bull Am Math Soc* 20(2):240–243
- Robert A (1985) Nonstandard analysis. *Bull Am Math Soc* 16(2):298–306
- Robinson A (1966) *Non-standard analysis*, rev edn. North Holland, Amsterdam
- Schmieden C, Laugwitz D (1958) Eine Erweiterung der Infinitesimalrechnung. *Math Zeit* 69:1–39
- van den Dries L, Wilkie AJ (1984) Gromov's theorem on groups of polynomial growth and elementary logic. *J Algebra* 89:349–374

Information System Design Using Fuzzy and Rough Set Theory

Frederick Petry
Naval Research Lab, Stennis Space Center,
Mississippi, MS, USA

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Fuzzy Sets and Rough Sets
Rough Relational Database
Information Theory
Rough Fuzzy Relational Database
Rough Set Modeling of Spatial Data
Data Mining in Rough Databases
Aggregation of Uncertain Information
Future Directions
Bibliography

Glossary

Rough Sets Rough set theory is a technique for dealing with uncertainty and for identifying cause-effect relationships in databases. It is based on a partitioning of some domain into equivalence classes and the defining of lower and upper approximation regions based on this partitioning to denote certain and possible inclusion in the rough set.

Fuzzy Sets Fuzzy set theory is another technique for dealing with uncertainty. It is based on the concept of measuring the degree of inclusion in a set through the use of a membership value. Where elements can either belong or not belong to a regular set, with fuzzy sets elements can belong to the set to a certain degree with zero indicating

not an element, one indicating complete membership, and values between zero and one indicating partial or uncertain membership in the set.

Information Theory Information theory involves the study of measuring the information content of a signal. In databases information theoretic measures can be used to measure the information content of data. Entropy is one such measure.

Database A collection of data and the application programs that make use of this data for some enterprise is a database.

Information System An information system is a database enhanced with additional tools that can be used by management for planning and decision-making.

Data Mining Data mining involves the discovery of patterns or rules in a set of data. These patterns generate some knowledge and information from the raw data that can be used for making decisions. There are many approaches to data mining, and uncertainty management techniques play a vital role in knowledge discovery.

Definition of the Subject and Its Importance

Databases and information systems are ubiquitous in this age of information and technology. Computers have revolutionized the way data can be manipulated and stored, allowing for very large databases with sophisticated capabilities. With so much money and manpower invested in the design and daily use of these systems, it is imperative that they be as correct, secure, and adaptable to the changing needs of the enterprise as possible. Therefore it is important to understand the design and implementation of such systems and to be able to utilize all their capabilities.

Scientists and business executives alike know the value of information. The challenge has been

to produce relevant information for an ever-changing uncertain world from data and facts stored on computers and archival devices. These data are considered to be exact, certain, factual values. The real world, however, is uncertain, inexact, and fraught with errors. It is a challenge, then, to extract useful and relevant information from ordinary databases. Uncertainty management techniques such as rough and fuzzy sets can help. These are emerging topics of importance in both the areas of data science/big data (Cady 2017; Dhar 2013) and the Internet of things (Höller et al. 2014; Stankovic 2014).

Introduction

Databases are recognized for their ability to store and update data in an efficient manner, providing reliability and the elimination of data redundancy. The relational database model, in particular, has well-established mechanisms built into the model for properly designing the database and maintaining data integrity and consistency. Data alone, however, are only facts. What is needed is information. Knowledge discovery attempts to derive information from the pure facts, discovering high-level regularities in the data. It is defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley et al. 1991; Han et al. 1992).

An innovative technique in the field of uncertainty and knowledge discovery is based on rough sets. Rough set theory, introduced and further developed mathematically by Pawlak (1984), provides a framework for the representation of uncertainty. It has been used in various applications such as the rough querying of crisp data (Beaubouef and Petry 1994b), uncertainty management in databases (Beaubouef and Petry 2007a), the mining of spatial data (Beaubouef and Petry 2002), and improved information retrieval (Srinivasan 1991). These techniques may readily be extended for use with object-oriented, spatial, and other complex databases and may be integrated with additional data mining

techniques for a comprehensive knowledge discovery approach.

Fuzzy Sets and Rough Sets

Fuzzy Set Theory

Fuzzy set theory (Zadeh 1965) is another approach for managing uncertainty. It has been around for a few years longer than rough sets and also has well-developed theory, properties, and applications. Applications involving fuzzy logic are diverse and plentiful, ranging from fuzzy control systems in industry to fuzzy logic in databases (Buckles and Petry 1982a; Petry 1996).

Definition

Characteristic Functions: Conventionally we can specify a set C by its characteristic function, $\text{Char}_C(x)$. If U is the universal set from which values of C are taken, then we can represent C as

$$C = \{x | x \in U \text{ and } \text{Char}_C(x) = 1\}$$

This is the representation for a crisp or non-fuzzy set. For an ordinary set C , the characteristic function is of the form:

$$\text{Char}_C(x) : U \rightarrow \{0, 1\}$$

However for a fuzzy set A we have

$$\text{Char}_F(x) : U \rightarrow [0, 1]$$

That is, for a fuzzy set the characteristic function takes on all values between 0 and 1 and not just the discrete values of 0 or 1 representing the binary choice for membership in a conventional crisp set such as C . For a fuzzy set, the characteristic function is often called the membership function and denoted $\mu_F(x)$.

Definition

Support and α -Cuts

The support of a fuzzy set, A , is a subset of the universe set, U ,

$$\text{Supp}(A) = \{x | x \in U \text{ and } \mu_A(x) > 0\}$$

So a fuzzy set A can be written as

$$A = \{\mu_A(x)/x | x \in \text{Supp}(A)\}$$

which means that only those fuzzy elements whose membership function is greater than zero contribute to A .

A related concept to the support is α -cuts. The α -cut of a set is a nonfuzzy set of the universe whose elements have a membership function greater than or equal to some value α . $A_\alpha = \{x | \mu_A(x) \geq \alpha\}$ for $0 \leq \alpha \leq 1$. Notice that the α -cuts of a set are subsets of the support. The values of α can be chosen arbitrarily but are usually picked to select desired subsets of the universe.

All of the basic set operations must have equivalent ones in fuzzy sets, but there are additional operations based on membership values of a fuzzy set that hence have no correspondence in crisp sets. We will use the membership functions μ_A and μ_B to represent the fuzzy sets A and B involved in the operations to be illustrated.

$$\text{Set Equality } A = B \Leftrightarrow \mu_A(x) = \mu_B(x)$$

$$\text{Set Containment } A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x)$$

$$\text{Set Complement } A = \{x/(1 - \mu_A(x))\}$$

For ordinary crisp sets $A \cap A = \emptyset$; however, this is not generally true for a fuzzy set and its complement. This may seem to violate the law of the excluded middle, but this is just the essential nature of fuzzy sets. Since fuzzy sets have imprecise boundaries, we cannot place an element exclusively in a set or its complement.

Set Union and Set Intersection

$$A \cup B \Leftrightarrow \mu_{A \cup B}(x) = \text{Max}(\mu_A(x), \mu_B(x))$$

$$A \cap B \Leftrightarrow \mu_{A \cap B}(x) = \text{Min}(\mu_A(x), \mu_B(x))$$

The justification for using the Max and Min functions for these operations is given in (28). With these definitions, the standard properties for crisp sets of commutativity, associativity, and

so forth hold for fuzzy sets. There have been a number of alternative functions proposed to represent set union and intersection (Cady 2017; Dubois and Prade 1992; Mendel 2017). For example, in the case of intersection, a product definition, $\mu_A(x) * \mu_B(x)$, has been considered.

$$\begin{aligned} \text{Concentration } \text{CON}(A) : \mu_{\text{CON}(A)}(x) \\ = (\mu_A(x))^2 \end{aligned}$$

The concentration operation concentrates fuzzy elements by reducing the membership grade proportionally more for elements that have smaller membership grades. This operation and the following ones of DIL and INT have no counterpart in ordinary set operations and are commonly used to represent linguistic hedges.

$$\text{Dilation } \text{DIL}(A) : \mu_{\text{DIL}(A)}(x) = (\mu_A(x))^{1/2}$$

The dilation operation dilates fuzzy elements by increasing the membership grade more for the elements with smaller membership grades.

Intensification (INT (A)). The intensification operation is like contrast intensification of a picture. It raises the membership grade of those elements within the crossover points and reduces the membership grade of those outside the crossover points.

The operators such as CON and DIL are commonly used to represent linguistic hedges that act as modifiers to linguistic variables represented in fuzzy sets. The CON operator can be used to approximate the effect of the linguistic modifier hedge “Very.” That is,

$$\text{Very}(A) = \text{CON}(A)$$

Another example is that the dilation operation may be used to represent the linguistic modifier More-or-less.

The exponents such as μ^2 for CON or $\mu^{1/2}$ for DIL can be viewed as specific values of parameterized exponents. So the hedge Extremely might be modeled by μ^3 or other values that might be obtained by a consensus of opinions.

Rough Set Theory

Rough set theory, introduced by Pawlak (1982, 1991), is a technique for dealing with uncertainty and for identifying cause-effect relationships in databases. An extensive theory for rough sets and their properties has been developed, and they have become a well-established approach for the management of uncertainty in a variety of applications. Rough sets involve the following:

U is the *universe*, which cannot be empty.

R is the *indiscernibility relation* or equivalence relation.

$A = (U, R)$, an ordered pair, is called an *approximation space*.

$[x]_R$ denotes the equivalence class of R containing x , for any element x of U .

Elementary sets in A – the equivalence classes of R .

Definable set in A – any finite union of elementary sets in A .

Given an approximation space defined on some universe U that has an equivalence relation R imposed upon it, U is partitioned into equivalence classes called elementary sets that may be used to define other sets in A . A rough set X , where $X \subseteq U$, can be defined in terms of the definable sets in A by the following:

Lower approximation of X in A is the set $\underline{R}X = \{x \in U / [x]_R \subseteq X\}$.

Upper approximation of X in A is the set $\overline{R}X = \{x \in U / [x]_R \cap X \neq \emptyset\}$.

$\text{POS}_R(X) = \underline{R}X$ denotes the R -positive region of X or those elements which certainly belong to the rough set. The R -negative region of X , $\text{NEG}_R(X) = U - \overline{R}X$, contains elements which do not belong to the rough set, and the boundary or R -borderline region of X , $\text{BN}_R(X) = \overline{R}X - \underline{R}X$, contains those elements which may or may not belong to the set. X is R -definable if and only if $\underline{R}X = \overline{R}X$. Otherwise, $\underline{R}X \neq \overline{R}X$ and X is rough with respect to R . A *rough set in A* is the group of subsets of U with the same upper and lower approximations.

Because there are advantages to both fuzzy set and rough set theories, several researchers have studied various ways of combining the two theories (Chanas and Kuchta 1992; Dubois and Prade 1987, 1992; Nanda and Majumdar 1992). Others have investigated the interrelations between the two theories (Chanas and Kuchta 1992; Pawlak 1985; Wygalak 1989). Fuzzy sets and rough sets are not equivalent, but complementary.

It has been shown in Wygalak (1989) that rough sets can be expressed by a fuzzy membership function $\mu \rightarrow \{0, 0.5, 1\}$ to represent the negative, boundary, and positive regions. In this model, all elements of the lower approximation, or positive region, have a membership value of one. Those elements of the boundary region are assigned a membership value of 0.5. Elements not belonging to the rough set have a membership value of zero. Rough set definitions of union and intersection can be modified so that the fuzzy model satisfies all the properties of rough sets (Beaubouef and Petry 2000).

We integrate fuzziness into the rough set model in order to quantify levels of roughness in boundary region areas through the use of fuzzy membership values. Therefore, we do not require membership values of elements of the boundary region to equal 0.5, but allow them to range from zero to one, noninclusive. Additionally, the union and intersection operators for fuzzy rough sets are comparable to those for ordinary fuzzy sets, where MIN and MAX are used to obtain membership values of redundant elements.

Let U be a *universe*, X a rough set in U .

Definition

A *fuzzy rough set* Y in U is a membership function $\mu_Y(x)$ which associates a grade of membership from the interval $[0,1]$ with every element of U where

$$\mu_Y(RX) = 1, \quad \mu_Y(U - \overline{R}X) = 0, \text{ and} \\ 0 < \mu_Y(\overline{R}X - RX) < 1.$$

Definition

The *union* of two fuzzy rough sets A and B is a fuzzy rough set C where

$$C = \{x | x \in A \text{ OR } x \in B\}, \text{ where } \mu_C(x) = \text{MAX}[\mu_A(x), \mu_B(x)].$$

Definition

The *intersection* of two fuzzy rough sets A and B is a fuzzy rough set C where

$$C = \{x | x \in A \text{ AND } x \in B\}, \text{ where } \mu_C(x) = \text{MIN}[\mu_A(x), \mu_B(x)].$$

Rough Relational Database

The rough relational database model (Beaubouef et al. 1995) is an extension of the standard relational database model of Codd (Petry 1996). It captures all the essential features of rough sets theory including indiscernibility of elements denoted by equivalence classes and lower and upper approximation regions for defining sets which are indefinable in terms of the indiscernibility.

Every attribute domain is partitioned by some equivalence relation designated by the database designer or user. Within each domain, those values that are considered indiscernible belong to an equivalence class. This information is used by the query mechanism to retrieve information based on equivalence with the class to which the value belongs rather than equality, resulting in less critical wording of queries.

Recall is also improved in the rough relational database because rough relations provide *possible* matches to the query in addition to the *certain* matches which are obtained in the standard relational database. This is accomplished by using set containment in addition to equality of attributes in the calculation of lower and upper approximation regions of the query result.

The rough relational database has several features in common with the ordinary relational database. Both models represent data as a collection of *relations* containing *tuples*. These relations are sets. The tuples of a relation are its elements and, like elements of sets in general, are unordered and nonduplicated. A tuple $\mathbf{t}_i$ takes the form $(d_{i1}, d_{i2}, \dots, d_{im})$, where d_{ij} is a *domain value* of a particular *domain set* D_j . In the ordinary relational

database, $d_{ij} \in D_j$. In the rough database, however, as in other non-first normal form extensions to the relational model (Makinouchi 1977; Roth et al. 1987), $d_{ij} \subseteq D_j$, and although it is not required that d_{ij} be a singleton, $d_{ij} \neq \emptyset$. Let $P(D_i)$ denote the powerset(D_i) – $\emptyset$.

Definition

A *rough relation* R is a subset of the set cross product $P(D_1) \times P(D_2) \times \dots \times P(D_m)$.

A rough tuple $\mathbf{t}$ is any member of R , which implies that it is also a member of $P(D_1) \times P(D_2) \times \dots \times P(D_m)$. If $\mathbf{t}_i$ is some arbitrary tuple, then $\mathbf{t}_i = (d_{i1}, d_{i2}, \dots, d_{im})$ where $d_{ij} \subseteq D_j$. A tuple in this model differs from that of ordinary databases in that the tuple components may be sets of domain values rather than single values. The set braces are omitted from singletons for notational simplicity.

Let $[d_{xy}]$ denote the equivalence class to which d_{xy} belongs. When d_{xy} is a set of values, the equivalence class is formed by taking the union of equivalence classes of members of the set; if $d_{xy} = \{c_1, c_2, \dots, c_n\}$, then $[d_{xy}] = [c_1] \cup [c_2] \cup \dots \cup [c_n]$.

Definition

Tuples $\mathbf{t}_i = (d_{i1}, d_{i2}, \dots, d_{im})$ and $\mathbf{t}_k = (d_{k1}, d_{k2}, \dots, d_{km})$ are *redundant* if $[d_{ij}] = [d_{kj}]$ for all $j = 1, \dots, m$.

In the rough relational database, redundant tuples are removed in the merging process since duplicates are not allowed in sets, the structure upon which the relational model is based.

There are two basic types of relational operators. The first type arises from the fact that relations are considered sets of tuples. Therefore, operations which can be applied to sets also apply to relations. The most useful of these for database purposes are *set difference*, *union*, and *intersection*. Operators which do not come from set theory, but which are useful for retrieval of relational data, are *select*, *project*, and *join*.

In the rough relational database, relations are rough sets as opposed to ordinary sets. Therefore,

new rough operators (--- , $\cup$, $\cap$, σ , π , $\bowtie$), which are comparable to the standard relational operators, must be developed for the rough relational database. Moreover, a mechanism must exist within the database to mark tuples of a rough relation as belonging to the lower or upper approximation of that rough relation. Properties of the rough relational operators can be found in Beaubouef et al. (1995).

Information Theory

In communication theory, Shannon (1948) introduced the concept of entropy which was used to characterize the information content of signals. Since then, variations of these information theoretic measures have been successfully applied to applications in many diverse fields. In particular, the representation of uncertain information by entropy measures has been applied to all areas of databases, including fuzzy database querying (Buckles and Petry 1983), data allocation (Fung and Lam 1980), classification in rule-based systems (Quinlan 1986), and measuring uncertainty in rough and fuzzy rough relational databases (Beaubouef et al. 1998).

In fuzzy set theory, the representation of uncertain information measures has been extensively studied (Bhandari and Pal 1993; de Luca and Termini 1972; Klir and Folger 1988). So this paper relates the concepts of information theory to rough sets and compares these information theoretic measures to established rough set metrics of uncertainty. The measures are then applied to the rough relational database model (Beaubouef et al. 1995). Information content of both stored relational schemas and rough relations are expressed as types of rough entropy.

Rough set theory (Pawlak 1982) inherently models two types of uncertainty. The first type of uncertainty arises from the indiscernibility relation that is imposed on the universe, partitioning all values into a finite set of equivalence classes. If every equivalence class contains only one value, then there is no loss of

information caused by the partitioning. In any coarser partitioning, however, there are fewer classes, and each class will contain a larger number of members. Our knowledge, or information, about a particular value decreases as the granularity of the partitioning becomes coarser.

Uncertainty is also modeled through the approximation regions of rough sets where elements of the lower approximation region have total participation in the rough set and those of the upper approximation region have uncertain participation in the rough set. Equivalently, the lower approximation is the *certain* region, and the boundary area of the upper approximation region is the *possible* region.

Pawlak (1991) discusses two numerical characterizations of imprecision of a rough set X : *accuracy* and *roughness*. Accuracy, which is simply the ratio of the number of elements in the lower approximation of X , $\underline{R}X$, to the number of elements in the upper approximation of the rough set X , $\overline{R}X$, measures the degree of completeness of knowledge about the given rough set X . It is defined as a ratio of the two set cardinalities as follows:

$$\alpha_R(X) = \text{card}(RX) / \text{card}(\overline{R}X), \text{ where} \\ 0 \leq \alpha_R(X) \leq 1.$$

The second measure, roughness, represents the degree of incompleteness of knowledge about the rough set. It is calculated by subtracting the accuracy from 1: $\rho_R(X) = 1 - \alpha_R(X)$.

These measures require knowledge of the number of elements in each of the approximation regions and are good metrics for uncertainty as it arises from the boundary region, implicitly taking into account equivalence classes as they belong wholly or partially to the set. However, accuracy and roughness measures do not necessarily provide us with information on the uncertainty related to the granularity of the indiscernibility relation for those values that are totally included in the lower approximation region. For example,

Let the rough set X be defined as follows:
 $X = \{A11, A12, A21, A22, B11, C1\}$
 with lower and upper approximation regions defined as

$$RX = \{A11, A12, A21, A22\} \text{ and}$$

$$\bar{R}X = \{A11, A12, A21, A22, B11, B12, B13, C1, C2\}$$

These approximation regions may result from one of several partitionings. Consider, for example, the following indiscernibility relations:

$$A_1 = \{[A11, A12, A21, A22], [B11, B12, B13], [C1, C2]\},$$

$$A_2 = \{[A11, A12], [A21, A22], [B11, B12, B13], [C1, C2]\},$$

$$A_3 = \{[A11], [A12], [A21], [A22], [B11, B12, B13], [C1, C2]\}.$$

All three of the above partitionings result in the same upper and lower approximation regions for the given set X and hence the same accuracy measure ($4/9 = 0.444$) since only those classes belonging to the lower approximation region were repartitioned. It is obvious, however, that there is more uncertainty in A_1 than in A_2 and more uncertainty in A_2 than in A_3 . Therefore, a more comprehensive measure of uncertainty is needed.

We derive such a measure from techniques used for measuring entropy in classical information theory. Countless variations of the classical entropy have been developed, each tailored for a particular application domain or for measuring a particular type of uncertainty. Our rough entropy is defined such that we may apply it to rough databases. We define the entropy of a rough set X as follows:

Definition

The *rough entropy* $E_r(X)$ of a rough set X is calculated by

$$E_r(x) = -(\rho_R(x)) \left[\sum Q_i \log(P_i) \right] \text{ for } i = 1, \dots, n \text{ equivalence classes.}$$

The term $\rho_R(X)$ denotes the roughness of the set X . The second term is the summation of the probabilities for each equivalence class belonging either wholly or in part to the rough set X . There is no ordering associated with individual class members. Therefore the probability of any one value of the class being named is the reciprocal of the number of elements in the class. If c_i is the cardinality of, or number of elements in, equivalence class i and all members of a given equivalence class are equal, $P_i = 1/c_i$ represents the probability of one of the values in class i . Q_i denotes the probability of equivalence class i within the universe. Q_i is computed by taking the number of elements in class i and dividing by the total number of elements in all equivalence classes combined. The entropy of the sample rough set X , $E_r(X)$, is given below for each of the possible indiscernibility relations A_1 , A_2 , and A_3 .

$$\text{Using } A_1: -(5/9)[(4/9)\log(1/4) + (3/9)\log(1/3) + (2/9)\log(1/2)] = 0.274$$

$$\text{Using } A_2: -(5/9)[(2/9)\log(1/2) + (2/9)\log(1/2) + (3/9)\log(1/3) + (2/9)\log(1/2)] = 0.20$$

$$\text{Using } A_3: -(5/9)[(1/9)\log(1) + (1/9)\log(1) + (1/9)\log(1) + (1/9)\log(1) + (3/9)\log(1/3) + (2/9)\log(1/2)] = 0.048$$

From the above calculations, it is clear that although each of the partitionings results in identical roughness measures, the entropy decreases as the classes become smaller through finer partitionings.

Entropy and the Rough Relational Database

The basic concepts of rough sets and their information-theoretic measures carry over to the rough relational database model (Beaubouef et al. 1995). Recall that in the rough relational database, all domains are partitioned into equivalence classes and relations are not restricted to first normal form. We therefore have a type of rough set for each attribute of a relation. This results in a rough relation, since any tuple having a value for an attribute that belongs to the boundary region of its domain is a tuple belonging to the boundary region of the rough relation.

There are two things to consider when measuring uncertainty in databases: uncertainty or entropy of a rough relation that exists in a database

at some given time and the entropy of a relation schema for an existing relation or query result. We must consider both since the approximation regions only come about by set values for attributes in given tuples. Without the extension of a database containing actual values, we only know about indiscernibility of attributes. We cannot consider the approximation regions.

We define the entropy for a rough relation schema as follows:

Definition

The *rough schema entropy* for a rough relation schema S is

$$E_s(S) = -\sum_j \left[\sum_i Q_i \log(P_i) \right] \quad \text{for} \\ i = 1, \dots, n; j = 1, \dots, m$$

where there are n equivalence classes of domain j , and m attributes in the schema $R(A_1, A_2, \dots, A_m)$.

This is similar to the definition of entropy for rough sets without factoring in roughness since there are no elements in the boundary region (lower approximation = upper approximation). However, because a relation is a cross product among the domains, we must take the sum of all these entropies to obtain the entropy of the schema. The schema entropy provides a measure of the uncertainty inherent in the definition of the rough relation schema taking into account the partitioning of the domains on which the attributes of the schema are defined.

We extend the schema entropy $E_s(S)$ to define the entropy of an actual rough relation instance $E_R(R)$ of some database D by multiplying each term in the product by the roughness of the rough set of values for the domain of that given attribute.

Definition

The *rough relation entropy* of a particular extension of a schema is

$$E_R(R) = -\sum_j D\rho_j(R) \left[\sum_i DQ_i \log(DP_i) \right] \quad \text{for} \\ i = 1, \dots, n; j = 1, \dots, m$$

where $D\rho_j(R)$ represents a type of database roughness for the rough set of values of the domain for attribute j of the relation, m is the number of attributes in the database relation, and n is the number of equivalence classes for a given domain for the database.

We obtain the $D\rho_j(R)$ values by letting the non-singleton domain values represent elements of the boundary region, computing the original rough set accuracy and subtracting it from one to obtain the roughness. DQ_i is the probability of a tuple in the database relation having a value from class i , and DP_i is the probability of a value for class i occurring in the database relation out of all the values which are given.

Information theoretic measures again prove to be a useful metric for quantifying information content. In rough sets and the rough relational database, this is especially useful since in ordinary rough sets, Pawlak's measure of roughness does not seem to capture the information content as precisely as our rough entropy measure.

In rough relational databases, knowledge about entropy can either guide the database user toward less uncertain data or act as a measure of the uncertainty of a data set or relation. As rough relations become larger in terms of the number of tuples or attributes, the automatic calculation of some measure of entropy becomes a necessity. Our rough relation entropy measure fulfills this need.

Rough Fuzzy Relational Database

The fuzzy rough relational database, as in the ordinary relational database, represents data as a collection of *relations* containing *tuples*. Because a relation is considered a set having the tuples as its members, the tuples are unordered. In addition, there can be no duplicate tuples in a relation. A tuple $\mathbf{t}_i$ takes the form $(d_{i1}, d_{i2}, \dots, d_{im}, d_{i\mu})$, where d_{ij} is a *domain value* of a particular *domain set* D_j and $d_{i\mu} \in D_\mu$, where D_μ is the interval $[0, 1]$, the domain for fuzzy membership values. In the ordinary relational database, $d_{ij} \in D_j$. In the fuzzy rough relational database, except for the fuzzy

membership value, however, $d_{ij} \subseteq D_j$, and although d_{ij} is not restricted to be a singleton, $d_{ij} \neq \emptyset$. Let $P(D_i)$ denote any non-null member of the powerset of D_i .

Definition

A *fuzzy rough relation* R is a subset of the set cross product $P(D_1) \times P(D_2) \times \cdots \times P(D_m) \times D_\mu$.

For a specific relation, R , membership is determined semantically. Given that D_1 is the set of names of nuclear/chemical plants, D_2 is the set of locations, and assuming that RIVERB is the only nuclear power plant that is located in VENTRESS,

```
(RIVERB, VENTRESS, 1)
(RIVERB, OSCAR, .7)
(RIVERB, ADDIS, 1)
(CHEMO, VENTRESS, .3)
```

are all elements of $P(D_1) \times P(D_2) \times D_\mu$. However, only the element (RIVERB, VENTRESS, 1) of those listed above is a member of the relation R (PLANT, LOCATION, μ), which associates each plant with the town or community in which it is located. A *fuzzy rough tuple* $\mathbf{t}$ is any member of R . If $\mathbf{t}_i$ is some arbitrary tuple, then $\mathbf{t}_i = (d_{i1}, d_{i2}, \dots, d_{im}, d_{i\mu})$ where $d_{ij} \subseteq D_j$ and $d_{i\mu} \in D_\mu$.

Definition

An *interpretation* $\alpha = (a_1, a_2, \dots, a_m, a_\mu)$ of a fuzzy rough tuple $\mathbf{t}_i = (d_{i1}, d_{i2}, \dots, d_{im}, d_{i\mu})$ is any value assignment such that $a_j \in d_{ij}$ for all j .

The interpretation space is the cross product $D_1 \times D_2 \times \cdots \times D_m \times D_\mu$, but is limited for a given relation R to the set of those tuples which are valid according to the underlying semantics of R . In an ordinary relational database, because domain values are atomic, there is only one possible interpretation for each tuple $\mathbf{t}_i$. Moreover, the interpretation of $\mathbf{t}_i$ is equivalent to the tuple $\mathbf{t}_i$. In the fuzzy rough relational database, this is not always the case.

Let $[d_{xy}]$ denote the equivalence class to which d_{xy} belongs. When d_{xy} is a set of values, the equivalence class is formed by taking the union of equivalence classes of members of the set; if

$d_{xy} = \{c_1, c_2, \dots, c_n\}$, then $[d_{xy}] = [c_1] \cup [c_2] \cup \dots \cup [c_n]$.

Definition

Tuples $\mathbf{t}_i = (d_{i1}, d_{i2}, \dots, d_{in}, d_{i\mu})$ and $\mathbf{t}_k = (d_{k1}, d_{k2}, \dots, d_{kn}, d_{k\mu})$ are *redundant* if $[d_{ij}] = [d_{kj}]$ for all $j = 1, \dots, n$.

If a relation contains only those tuples of a lower approximation, i.e., those tuples having a μ value equal to one, the interpretation α of a tuple is unique. This follows immediately from the definition of redundancy. In fuzzy rough relations, there are no redundant tuples. The merging process used in relational database operations removes duplicate tuples since duplicates are not allowed in sets, the structure upon which the relational model is based.

Tuples may be redundant in all values except μ . As in the union of fuzzy rough sets where the maximum membership value of an element is retained, it is the convention of the fuzzy rough relational database to retain the tuple having the higher μ value when removing redundant tuples during merging. If we are supplied with identical data from two sources, one certain and the other uncertain, we would want to retain the data that is certain, avoiding loss of information.

Recall that the rough relational database is in non-first normal form; there are some attribute values that are sets. Another definition, which will be used for upper approximation tuples, is necessary for some of the alternate definitions of operators to be presented. This definition captures redundancy between elements of attribute values that are sets:

Definition

Two sub-tuples $X = (d_{x1}, d_{x2}, \dots, d_{xm})$ and $Y = (d_{y1}, d_{y2}, \dots, d_{ym})$ are *roughly redundant*, $\approx_R$, if for some $[p] \subseteq [d_{xj}]$ and $[q] \subseteq [d_{yj}]$, $[p] = [q]$ for all $j = 1, \dots, m$.

In order for any database to be useful, a mechanism for operating on the basic elements and retrieving specified data must be provided. The concepts of redundancy and merging play a key role in the operations defined.

We must first design our database using some type of semantic model. We use a variation of the entity-relationship diagram that we call a fuzzy rough E-R diagram. This diagram is similar to the standard E-R diagram in that entity types are depicted in rectangles, relationships with diamonds, and attributes with ovals. However, in the fuzzy rough model, it is understood that membership values exist for all instances of entity types and relationships. Attributes which allow values where we want to be able to define equivalences are denoted with an asterisk (*) above the oval. These values are defined in the indiscernibility relation, which is not actually part of the database design, but inherent in the fuzzy rough model.

Our fuzzy rough E-R model (Beaubouef and Petry 2000) is similar to the second and third levels of fuzziness defined by Zvieli and Chen (1986). However, in our model, all entity and relationship occurrences (second level) are of the fuzzy type so we do not mark an “F” beside each one. Zvieli and Chen’s third level considers attributes that may be fuzzy. They use triangles instead of ovals to represent these attributes. We do not introduce fuzziness at the attribute level of our model in this paper, only roughness or indiscernibility, and denote those attribute with the “*.” From the fuzzy rough E-R diagram, we design the structure of the fuzzy rough relational database. If we have a priori information about the types of queries that will be involved, we can make intelligent choices that will maximize computer resources.

We next formally define the fuzzy rough relational database operators and discuss issues relating to the real-world problems of data representation and modeling. We may view indiscernibility as being modeled through the use of the indiscernibility relation, imprecision through the use of non-first normal form constructs, and degree of uncertainty and fuzziness through the use of tuple membership values, which are given as the value for the μ attribute in every fuzzy rough relation.

Fuzzy Rough Relational Operators

In Beaubouef et al. (1995), we defined several operators for the rough relational algebra. We

now define similar operators for the fuzzy rough relational database as in Beaubouef and Petry (1994a). Recall that for all of these operators, the indiscernibility relation is used for equivalence of attribute values rather than equality of values.

Difference

The fuzzy rough relational difference operator is very much like the ordinary difference operator in relational databases and in sets in general. It is a binary operator that returns those elements of the first argument that are not contained in the second argument.

In the fuzzy rough relational database, the difference operator is applied to two fuzzy rough relations and, as in the rough relational database, indiscernibility, rather than equality of attribute values, is used in the elimination of redundant tuples. Hence, the difference operator is somewhat more complex. Let X and Y be two union compatible fuzzy rough relations.

Definition

The *fuzzy rough difference*, $X - Y$, between X and Y is a fuzzy rough relation T where

$$T = \{t(d_1, \dots, d_n, \mu_i) \in X \mid t(d_1, \dots, d_n, \mu_i) \notin Y\} \cup \{t(d_1, \dots, d_n, \mu_i) \in X \mid t(d_1, \dots, d_n, \mu_j) \in Y \text{ and } \mu_i > \mu_j\}$$

The resulting fuzzy rough relation contains all those tuples which are in the lower approximation of X , but not redundant with a tuple in the lower approximation of Y . It also contains those tuples belonging to upper approximation regions of both X and Y , but which have a higher μ value in X than in Y . For example, let X contain the tuple (**MODERN**, 1) and Y contain the tuple (**MODERN**, .02). It would not be desirable to subtract out certain information with possible information, so $X - Y$ yields (**MODERN**, 1).

Union

Because relations in databases are considered as sets, the union operator can be applied to any two union-compatible relations to result in a third relation which has as its tuples all the tuples contained in either or both of the two original relations. The

union operator can be extended to apply to fuzzy rough relations. Let X and Y be two union compatible fuzzy rough relations.

Definition

The *fuzzy rough union* of X and Y , $X \cup Y$ is a fuzzy rough relation T where

$$T = \{t | t \in X \text{ OR } t \in Y\} \quad \text{and} \quad \mu_T(t) = \text{MAX}[\mu_X(t), \mu_Y(t)].$$

The resulting relation T contains all tuples in either X or Y or both, merged together and having redundant tuples removed. If X contains a tuple that is redundant with a tuple in Y except for the μ value, the merging process will retain only that tuple with the higher μ value.

Intersection

The fuzzy rough intersection, another binary operator on fuzzy rough relations, can be defined similarly.

Definition

The *fuzzy rough intersection* of X and Y , $X \cap Y$ is a fuzzy rough relation T where

$$T = \{t | t \in X \text{ AND } t \in Y\} \quad \text{and} \quad \mu_T(t) = \text{MIN}[\mu_X(t), \mu_Y(t)].$$

In intersection, the MIN operator is used in the merging of equivalent tuples having different μ values, and the result contains all tuples that are members of both of the original fuzzy rough relations.

Definition

The *fuzzy rough intersection* of X and Y , $X \cap_A Y$ is a fuzzy rough relation T where

$$T = \{t | t \in X, \text{ and } \exists s \in Y | t \approx_R s\} \cup \{s | s \in Y, \text{ and } \exists t \in X | s \approx_R t\} \text{ and}$$

$$\mu_T(t) = \text{MIN}[\mu_X(t), \mu_Y(t)].$$

Select

The select operator for the fuzzy rough relational database model, σ , is a unary operator which takes a fuzzy rough relation X as its argument and returns a fuzzy rough relation containing a subset of the tuples of X , selected on the basis of values for a specified attribute. The operation $\sigma_{A=a}(X)$, for example, returns those tuples in X where attribute A is equivalent to the class $[a]$. In general, select returns a subset of the tuples that match some selection criteria.

Let R be a relation schema, X a fuzzy rough relation on that schema, and A an attribute in R , $\mathbf{a} = \{a_i\}$ and $\mathbf{b} = \{b_j\}$, where $a_i, b_j \in \text{dom}(A)$ and $\cup_x$ is interpreted as “the union over all x .”

Definition

The *fuzzy rough selection*, $\sigma_{A=a}(X)$, of tuples from X is a fuzzy rough relation Y having the same schema as X and where

$$Y = \{t \in X | \cup_i [a_i] \subseteq \cup_j [b_j]\},$$

and $a_i \in \mathbf{a}$, $b_j \in t(A)$, and where membership values for tuples are calculated by multiplying the original membership value by

$$\text{card}(\mathbf{a})/\text{card}(\mathbf{b})$$

where $\text{card}(x)$ returns the cardinality, or number of elements, in x .

Assume we want to retrieve those elements where **CITY** = “ADDIS” from the following fuzzy rough tuples:

(ADDIS	1)
({ADDIS, LOTTIE, BRUSLY}	.7)
(OSCAR	1)
({ADDIS, JACKSON}	.9)

The result of the selection is the following:

(ADDIS	1)
({ADDIS, LOTTIE, BRUSLY}	.23)
({ADDIS, JACKSON}	.45)

where the μ for the second tuple is the product of the original membership value 0.7 and $1/3$.

Project

Project is a unary fuzzy rough relational operator. It returns a relation that contains a subset of the columns of the original relation. Let X be a fuzzy rough relation with schema A , and let B be a subset of A . The fuzzy rough projection of X onto B is a fuzzy rough relation Y obtained by omitting the columns of X which correspond to attributes in $A - B$, and removing redundant tuples. Recall the definition of redundancy accounts for indiscernibility, which is central to the rough sets theory and that higher μ values have priority over lower ones.

Definition

The *fuzzy rough projection* of X onto B , $\pi_B(X)$, is a fuzzy rough relation Y with schema $Y(B)$ where

$$Y(B) = \{t(B) | t \in X\}.$$

Join

Join is a binary operator that takes related tuples from two relations and combines them into single tuples of the resulting relation. It uses common attributes to combine the two relations into one, usually larger, relation. Let $X(A_1, A_2, \dots, A_m)$ and $Y(B_1, B_2, \dots, B_n)$ be fuzzy rough relations with m and n attributes, respectively, and $AB = C$, the schema of the resulting fuzzy rough relation T .

Definition

The *fuzzy rough join*, $X \bowtie_{\langle \text{JOIN CONDITION} \rangle} Y$, of two relations X and Y , is a relation $T(C_1, C_2, \dots, C_{m+n})$ where

$T = \{t | \exists t_X \in X, t_Y \in Y \text{ for } t_X = t(A), t_Y = t(B)\}$, and where

$$t_X(A \cap B) = t_Y(A \cap B), \mu = 1 \quad (1)$$

$$\begin{aligned} t_X(A \cap B) &\subseteq t_Y(A \cap B) \text{ or } t_Y(A \cap B) \\ &\subseteq t_X(A \cap B), \mu = \text{MIN}(\mu_X, \mu_Y) \end{aligned} \quad (2)$$

$\langle \text{JOIN CONDITION} \rangle$ is a conjunction of one or more conditions of the form $A = B$.

Only those tuples which resulted from the “joining” of tuples that were both in lower approximations in the original relations belong to the lower approximation of the resulting fuzzy rough relation. All other “joined” tuples belong to the upper approximation only (the boundary region) and have membership values less than one. The fuzzy membership value of the resultant tuple is simply calculated as in Buckles and Petry (1985) by taking the minimum of the membership values of the original tuples. Taking the minimum value also follows the logic of Ola and Ozsoyoglu (1993), where, in joins of tuples with different levels of information uncertainty, the resultant tuple can have no greater certainty than that of its least certain component.

Fuzzy and rough set techniques integrated into the underlying data model result in databases that can more accurately represent real-world enterprises since they incorporate uncertainty management directly into the data model itself. This is useful as is for obtaining greater information through the querying of rough and fuzzy databases. Additional benefits may be realized when they are used in the process of data mining.

Rough Set Modeling of Spatial Data

Many of the problems associated with data are prevalent in all types of database systems. Spatial databases and GIS contain descriptive as well as positional data (Jing and Wenwen 2016). The various forms of uncertainty occur in both types of data, so many of the issues apply to ordinary databases as well, such as integration of data from multiple sources, time-variant data, uncertain data, imprecision in measurement, inconsistent wording of descriptive data, and “binning” or grouping of data into fixed categories, which also are employed in spatial contexts (Petry et al. 2005; Tavana et al. 2016).

First consider an example of the use of rough sets in representing spatially related data. Let $U = \{\text{tower, stream, creek, river, forest,}$

woodland, pasture, meadow}, and let the equivalence relation R be defined as follows:

$$R^* = \{[tower], [stream, creek, river], [forest, woodland], [pasture, meadow]\}.$$

Given some set $X = \{tower, stream, creek, river, forest, pasture\}$, we would like to define it in terms of its lower and upper approximations:

$$RX = \{tower, stream, creek, river\},$$

$$\bar{R}X = \{tower, stream, creek, river, forest, woodland, pasture, meadow\}.$$

A *rough set* in A is the group of subsets of U with the same upper and lower approximations. In the example given, the rough set is

$$\begin{aligned} &\{\{tower, stream, creek, river, forest, pasture\} \\ &\{tower, stream, creek, river, forest, meadow\} \\ &\{tower, stream, creek, river, woodland, pasture\} \\ &\{tower, stream, creek, river, woodland, meadow\}\}. \end{aligned}$$

Often spatial data is associated with a particular grid. The positions are set up in a regular matrix-like structure, and data is affiliated with point locations on the grid. This is the case for raster data and for other types of non-vector type data such as topography or sea surface temperature data. There is a trade-off between the resolution or the scale of the grid and the amount of system resources necessary to store and process the data. Higher resolutions provide more information, but at a cost of memory space and execution time.

If we approach these data issues from a rough set point of view, it can be seen that there is indiscernibility inherent in the process of gridding or rasterizing data. A data item at a particular grid point in essence may represent data near the point as well. This is due to the fact that often point data must be mapped to the grid using techniques such as nearest-neighbor, averaging, or statistics. The rough set indiscernibility relation may be set up so that the entire spatial area is partitioned into equivalence classes where each point on the grid

belongs to an equivalence class. If the resolution of the grid changes, then, in fact, this is changing the granularity of the partitioning, resulting in fewer, but larger classes.

The approximation regions of rough sets are beneficial whenever information concerning spatial data regions is accessed. Consider a region such as a forest. One can reasonably conclude that any grid point identified as FOREST that is surrounded on all sides by grid points also identified as FOREST is, in fact, a point represented by the feature FOREST. However, consider points identified as FOREST that are adjacent to points identified as MEADOW. Is it not possible that these points represent meadow area as well as forest area but were identified as FOREST in the classification process? Likewise, points identified as MEADOW but adjacent to FOREST points may represent areas that contain part of the forest. This uncertainty maps naturally to the use of the approximation regions of the rough set theory, where the lower approximation region represents certain data and the boundary region of the upper approximation represents uncertain data. It applies to spatial database querying and spatial database mining operations.

If we force a finer granulation of the partitioning, a smaller boundary region results. This occurs when the resolution is increased. As the partitioning becomes finer and finer, finally a point is reached where the boundary region is nonexistent. Then the upper and lower approximation regions are the same, and there is no uncertainty in the spatial data as can be determined by the representation of the model.

In Worboys (1998a) Worboys models imprecision in spatial data based on the resolution at which the data is represented and for issues related to the integration of such data. This approach relies on the issue of indiscernibility – a core concept for rough sets – but does not carry over the entire framework and is just described as “reminiscent of the theory of rough sets” (Worboys 1998b). Ahlqvist and colleagues (Ahlqvist et al. 2000) used a rough set approach to define a rough classification of spatial data and to represent spatial locations. They also proposed a measure for quality of a rough classification compared to a crisp classification and evaluated their technique on actual data from

vegetation map layers. They considered the combination of fuzzy and rough set approaches for reclassification as required by the integration of geographic data. Another research group in a mapping and GIS context (Wang et al. 2002) have developed an approach using a rough raster space for the field representation of a spatial entity and evaluated it on a classification case study for remote sensing images. In (2003) Bittner and Stell consider K-labeled partitions, which can represent maps, and then develop their relationship to rough sets to approximate map objects with vague boundaries. Additionally they investigate stratified partitions, which can be used to capture levels of details or granularity such as in consideration of scale transformations in maps, and extend this approach using the concepts of stratified rough sets. An additional approach for dealing with uncertain spatial data can be done using a Dempster-Shafer representation (Shafer 1976). By considering uncertainty in a spatial location as having a range that is most probable and an outer range that is possible, then by using nested intervals around a point, the uncertainty can be modeled (Elmore et al. 2017b).

Data Mining in Rough Databases

Association rules capture the idea of certain data items commonly occurring together and have been often considered in the analysis of a “market basket” of purchases. For example, a delicatessen retailer might analyze the previous year’s sales and observe that of all purchases 30% were of *both* cheese *and* crackers and, for any of the sales that included cheese, 75% also included crackers. Then it is possible to conclude a rule of the form:

$$\text{Cheese} \rightarrow \text{Crackers}$$

This rule is said to have a 75% degree of confidence and a 30% degree of support. This particular form of data is largely based on the Apriori algorithm developed by Agrawal et al. (1993). Let a database of possible data items be Han and Kamber (2006).

$$D = \{d_1, d_2, \dots, d_n\}$$

and the relevant set of transactions (sales, query results, etc.) are

$$T = \{T_1, T_2, \dots\}$$

where $T_i \subseteq D$. We are interested in discovering if there is a relationship between two sets of items (called itemsets) X_j, X_k ; $X_j, X_k \subseteq D$. For such a relationship to be determined, the entire set of transactions in T must be examined and a count made of the number of transactions containing these sets, where a transaction T_i contains X_m if $X_m \subseteq T_i$. This count, called the support count of X_m , $SC_T(X_m)$, will be appropriately modified in the case of rough sets.

There are then two measures used in determining rules: the percentage of T_i ’s in T that.

1. Contain both X_j and X_k (i.e., $X_j \cup X_k$) – called the *support* s .
2. If T_i contains X_j then T_i also contains X_k – called the *confidence* c .

The support and confidence can be interpreted as probabilities:

1. $s = \text{Prob}(X_j \cup X_k)$ and
2. $c = \text{Prob}(X_k | X_j)$

We assume the system user has provided minimum values for these in order to generate only sufficiently interesting rules. A rule whose support and confidence exceeds these minimums is called a strong rule.

The result of a query is then

$$T = \{RT, \overline{RT}\}$$

and so we must take into account the lower approximation, $\underline{RT}$, and upper approximation, $\overline{RT}$, results of the rough query in developing the association rules.

Recall that in order to generate frequent itemsets, we must count the number of transactions T_j that support an itemset X_j . In the ordinary

data mining algorithm, one simply counts the occurrence of a value as 1 if in the set or 0 if not. But now since the query result T is a rough set, we must modify the support count SC_T . So we define the rough support count, RSC_T , for the set X_j , to count differently in the upper and lower approximations:

$$RSC_T(X_j) = \sum_i W_{T_i}(X_j); X_j \subseteq T_i$$

$$\text{where } W(X_j) = \begin{cases} 1 & \text{if } T_i \in \underline{RT} \\ a & \text{if } T_i \in \overline{RT}, \quad 0 < a < 1 \end{cases}$$

The value, a , can be a subjective value obtained from the user depending on relative assessment of the roughness of the query result T . For the data mining example of the next section, we chose a neutral default value of $a = 1/2$. Note that $W(X_j)$ is included in the summation only if all of the values of the itemset X_j are included in the transaction, i.e., it is a subset of the transaction.

Finally to produce the association rules from the set of relevant data T retrieved from the spatial database, we must consider the frequent itemsets. For the purposes of generating a rule such as $X_j \rightarrow X_k$, we can now extend the approach to rough support and confidence as follows:

$$RS = RSC_T(X_j \cup X_k) / |T|$$

$$RC = RSC_T(X_j \cup X_k) / RSC_T(X_j)$$

In the spatial data mining area, there have only been a few efforts using rough sets. In the research described in (Beaubouef and Petry 2002; Bhattacharya and Bhatnagar 2012), approaches for attribute induction knowledge discovery (Raschia and Mouaddib 2002) in rough spatial data are investigated. In Bittner (2000) Bittner considers rough sets for spatiotemporal data and how to discover characteristic configurations of spatial objects focusing on the use of topological relationships for characterizations. In a survey of uncertainty-based spatial data mining, Shi et al. (2003) provide a brief general comparison of

fuzzy and rough set approaches for spatial data mining.

Aggregation of Uncertain Information

Uncertainty arising from multiple sources and of many forms appears in the everyday activities and decisions of humans. We want to examine approaches that can be used to combine these uncertainties into forms that can become useful for decision-making. Effective decision-making should be able to make use of all the available, relevant information about such combined uncertainty (Ferson and Kreinovich 2001). In this section we describe approaches for combining separately possibilistic uncertainty, probabilistic uncertainty, and situations where both forms of uncertainty appear.

To formalize the discussion, let V be a discrete variable taking values in a space X that has both aleatory and epistemic sources of uncertainty (Parsons 2001). Let P be a probability distribution $P: X \rightarrow [0, 1]$ such that $p_k \in [0, 1]$, $\sum_{k=1}^n p_k = 1$, that models the aleatory uncertainty. Then the epistemic uncertainty can be modeled by a possibility distribution (Makinouchi 1977) such that $\Pi: X \rightarrow [0, 1]$, where $\pi(x_k)$ gives the possibility that x_k is the value of V , $k = 1 \dots n$. A usual requirement here is the normality condition, $\max_x [\pi(x)] = 1$, that is at least one element in X must be fully possible. Abbreviating our notation so that $p_k = p(x_k)$, $\dots$ and $\pi_k = \pi(x_k)$, $\dots$, we have $P = \{p_1, p_2, \dots, p_n\}$ and $\Pi = \{\pi_1, \pi_2, \dots, \pi_n\}$.

Definition

Shannon Entropy

Shannon entropy has been the most broadly applied measure of randomness or information content (Shannon 1948). For a probability distribution $P = \{p_1, p_2, \dots, p_n\}$ this is given as

$$S(P) = - \sum_{i=1}^n p_i \ln(p_i).$$

Definition*Gini Index*

The Gini index, $G(P)$, also known as the Gini coefficient, is a measure of statistical dispersion developed by Gini (1912), and is defined as

$$G(P) \equiv 1 - \sum_{i=1}^n p_i^2$$

Some practitioners use $G(P)$ versus $S(P)$ since it does not involve a logarithm, making analytic solutions simpler. Gini index is used in consideration of inequalities in various areas such as economics, ecology, and engineering (Aristondo et al. 2012). A very important application of the Gini index is as a splitting criterion for decision tree induction in machine learning and data mining (Breiman et al. 1984).

It is accepted in practice for diagnostic test selection that the Shannon and Gini measures are interchangeable (Sent and van de Gaag 2007). The specific relationship of Shannon entropy and the Gini index has been discussed in the literature (Eliazar and Sokolov 2010). Theoretical support for this practice is provided in Yager's independent consideration of alternative measures of entropy (Yager 1995), where he derives the same form for an entropy measure as the Gini measure.

Definition*Specificity*

The concept of specificity in the framework of possibility theory has a function analogous to the entropy measures for probability we have discussed above. Yager (1982) gave a formal definition of the specificity of a fuzzy subset which was extended for possibility distributions as

$$Sp(\Pi) = \int_0^{\alpha_{\max}} \frac{1}{\text{card}(\Pi_\alpha)} d\alpha$$

where $\alpha_{\max} = \text{Max}_x \Pi(x)$ and is the α -possibility level of Π , the subset of elements having possibility of at least α . A linear specificity measure provides an intuitive measure of specificity of a possibility distribution (Pedrycz and Gomide

1996). Let $\pi_m = \text{Max}_{k=1}^n \pi_k$, and then the specificity measure is this max value minus the average of the other possibility values (Yager 1992):

$$Sp(\Pi) = \pi_m - \left(\sum_{k=1, k \neq m}^n \pi_k \right) / n$$

Definition*Consistency*

There have been a number of approaches to consistency measures of probability and possibility distributions that have been proposed, in particular the Zadeh (1978) consistency measure,

$$C_Z(P, \Pi) = \sum_{i=1}^n p_i * \pi_i$$

This measure does not represent an inherent relationship but rather represents the intuition that a lowering of an event's possibility tends to lower its probability, but not the converse.

There is some research to our specific concern of aggregating probabilistic and possibilistic uncertainty. First a related approach is the possibilistic conditioning of probability distributions using the approach of Yager (2012). This form of aggregation makes it very amenable to apply the information measures (Elmore et al. 2014).

In possibilistic conditioning, a function F dependent on both P and Π is used to find a new conditioned probability distribution such that

$$\hat{P} = F(P, \Pi)$$

where $\hat{P} = \{\hat{p}_1, \hat{p}_2, \dots, \hat{p}_n\}$ with

$$\hat{p}_n = p_i \pi_i / K; \quad K = \sum_{i=1}^n p_i * \pi_i$$

A strength of conditioned probability is that it also captures Zadeh's concept of consistency

between the possibility and the original probability distribution. Note that the K is of the form of the consistency expression:

$$C_Z(p, \Pi) = K.$$

Another approach uses transformation of possibility distributions into probability distributions. A number of possibility transformations have been considered (Beaubouef and Petry 2007b; Bhandari and Pal 1993; Elmore et al. 2017a), and here we will utilize one initially suggested by Dubois and Prade (1983). For this we must assume a possibility distribution where

$$\pi_1 \geq \pi_2 \geq \dots \geq \pi_n \text{ and } \pi_1 = 1.$$

Then we can obtain an associated probability distribution $P_1 = (p_{11}, \dots, p_{1n})$ (denoting $\pi_{n+1} = 0$) where

$$p_{1j} = \sum_{k=j}^n (\pi_k - \pi_{k+1})/k, \quad \text{for } j = 1 \text{ to } n$$

The distribution P_1 from the transformation can then be aggregated with some other probability distribution P_2 using a spectrum of operators such as min, max, mean, etc. (Petry et al. 2015). The aggregated probability distribution can then be evaluated using the Gini information measure to assess if there is enhanced information content over that of the initial distribution P_2 . Additional recent approaches have considered an intelligent quality-based approach to fusing multi-source probabilistic information (Yager and Petry 2016) and use of fuzzy Choquet integration of homogeneous possibility and probability distributions (Anderson et al. 2016).

Future Directions

There are several other approaches to uncertainty representation that may be more suitable for certain applications. Type 2 fuzzy sets have been of considerable recent interest (Mendel and John 2002). In these as opposed to ordinary fuzzy sets

in which the underlying membership functions are crisp, here the membership functions are themselves fuzzy. Intuitionistic sets introduced by Atanassov (1986, 2000) are another generalization of a fuzzy set. Two characteristic functions are used for capturing both the ordinary idea of degree of membership in the intuitionistic set and the degree of non-membership of elements in the set and can be used in database design (Beaubouef and Petry 2007b). Related to the concepts introduced by rough sets is the idea of granularity for managing complex data by abstraction using information granules as discussed by Lin (1997, 1999). A granular set approach has also been introduced (Ligeza 2002) which is a set and a number of disjoint subsets that constitute a semi-partition. Use of a Dempster-Shafer approach analogous to rough sets can be used to model uncertainty in spatial, temporal, and spatiotemporal application domains (Elmore et al. 2017a, b). Some prior database research on ordered relations (Ginsburg and Hull 1983), although not presented in the context of uncertainty of data, may provide some approaches to extend our work in this area. A main emphasis for future work is the incorporation of some of these research topics into mainstream database, GIS commercial products, and semi-structured data on the semantic web.

Acknowledgments The authors would like to thank the Naval Research Laboratory's Base Program, Program Element No. 0602435 N, for sponsoring this research.

Bibliography

Primary Literature

- Agrawal R, Imielinski T, Swami A (1993) Mining Association Rules between sets of items in large databases. Proceedings of the 1993 ACM-SIGMOD international conference on Management of Data. ACM Press, New York, pp 207–216
- Ahlqvist O, Keukelaar J, Oukbir K (2000) Rough classification and accuracy assessment. *Int J Geogr Inf Sci* 14:475–496
- Anderson, D, Elmore P, Petry, F Havens T (2016) Fuzzy Choquet integration of homogenous possibility and probability distributions. *Inf Sci* 363:24–39
- Aristondo O, Garcia-Lparesa J, de la Vega C, Pereira R (2012) The Gini index, the dual decomposition of

- aggregation functions and the consistent measurement of inequality. *Int J Intell Syst* 27:132–152
- Atanassov K (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20:87–96
- Atanassov K (2000) Intuitionistic fuzzy sets; theory and applications. Physica Verlag, Heidelberg
- Beaubouef T, Petry F (1994a) Fuzzy set quantification of roughness in a rough relational database model. In: *Proceedings of the third IEEE international conference on fuzzy systems*, Orlando, pp 172–177
- Beaubouef T, Petry F (1994b) Rough querying of crisp data in relational databases. In: *Proceedings of the third international workshop on rough sets and soft computing (RSSC'94)*, San Jose, Hershey, pp 368–375
- Beaubouef T, Petry F (2000) Fuzzy rough set techniques for uncertainty processing in a relational database. *Int J Intell Syst* 15:389–424
- Beaubouef T, Petry F (2002) A rough set foundation for spatial data mining involving vague regions. In: *Proceedings of FUZZ-IEEE'02*, Honolulu, pp 767–772
- Beaubouef T, Petry F (2007a) Rough sets: a versatile theory for approaches to uncertainty Management in Databases. *Rough Computing: Theories, Technologies and Applications*, Idea Group, Inc
- Beaubouef T, Petry F (2007b) Intuitionistic rough sets for database applications. In: *Peters JF et al (eds) Transactions on rough sets VI. LNCS 4374*. Springer, Berlin/New York, pp 26–30
- Beaubouef T, Petry F, Buckles B (1995) Extension of the relational database and its algebra with rough set techniques. *Comput Intell* 11:233–245
- Beaubouef T, Petry F, Arora G (1998) Information-theoretic measures of uncertainty for rough sets and rough relational databases. *Inf Sci* 109:185–195
- Bhandari D, Pal NR (1993) Some new information measures for fuzzy sets. *Inf Sci* 67:209–228
- Bhattacharya S, Bhatnagar V (2012) Fuzzy data mining: a literature survey and classification framework. *Int J Netw Virt Org* 11:382–408
- Bittner T (2000) Rough sets in spatio-temporal data mining. *Proceedings of international workshop on temporal, spatial and spatio-temporal data mining*. Springer, Berlin/Heidelberg, pp 89–104
- Bittner T, Stell J (2003) Stratified rough sets and vagueness. In: *Kuhn W, Worboys M, Timpf S (eds) Spatial information theory: cognitive and computational foundations of geographic information science international conference (COSIT'03)* pp 286–303
- Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and regression trees*. Wadsworth & Brooks/Cole, Monterey
- Buckles B, Petry F (1982) A fuzzy representation for relational data bases. *Int J Fuzzy Sets Syst* 7(3):213–226
- Buckles B, Petry F (1983) Information-theoretical characterization of fuzzy relational databases. *IEEE Trans Syst Man Cybern* 13:74–77
- Buckles BP, Petry F (1985) Uncertainty models in information and database systems. *J Inf Sci* 11:77–87
- Cady F (2017) *The data science handbook*. Wiley, New York
- Chanas S, Kuchta D (1992) Further remarks on the relation between rough and fuzzy sets. *Fuzzy Sets Syst* 47:391–394
- de Luca A, Termini S (1972) A definition of a non-probabilistic entropy in the setting of fuzzy set theory. *Inf Control* 20:301–312
- Dhar V (2013) Data science and prediction. *Commun ACM* 56(12):64–73
- Dubois D, Prade H (1983) Unfair coins and necessity measures: towards a possibilistic interpretations of histograms. *Fuzzy Sets Syst* 10:15–27
- Dubois D, Prade H (1987) Twofold fuzzy sets and rough sets—some issues in knowledge representation. *Fuzzy Sets Syst* 23:3–18
- Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: *Slowinski R (ed) Intelligent decision support: handbook of applications and advances of the rough sets theory*. Kluwer Academic Publishers, Boston
- Eliazar I, Sokolov I (2010) Maximization of statistical heterogeneity: from Shannon's entropy to Gini's index. *Phys A* 389:3023–3038
- Elmore P, Petry F, Yager R (2014) Comparative measures of aggregated uncertainty representations. *J Ambient Intell Humaniz Comput* 5(6):809–819
- Elmore P, Petry F, Yager R (2017a) Dempster-Shafer approach to temporal uncertainty. *IEEE Trans Emerg Topics Comput Intell* 1(5):316–325
- Elmore P, Petry F, Yager R (2017b) Geospatial modeling using dempster-shafer theory. *IEEE Trans Cybern* 47(6):1551–1561
- Ferson S, Kreinovich V (2001) Representation, elicitation, and aggregation of uncertainty in risk analysis – from traditional probabilistic techniques to more general, more realistic approaches: a survey. University of Texas at El Paso computer science tech report #11-1-2001
- Frawley W, Piatetsky-Shapiro G, Matheus C (1991) Knowledge discovery in databases: an overview. In: *Piatetsky-Shapiro G, Frawley W (eds.), Knowledge discovery in databases*, AAAI/MIT Press, Menlo Park pp 1–27
- Fung KT, Lam CM (1980) The database entropy concept and its application to the data allocation problem. *Infor* 18(4):354–363
- Gini C (1912) *Variabilita e mutabilita (Variability and Mutability)*. Tipografia di Paolo Cuppini, Bologna, p 156
- Ginsburg S, Hull R (1983) Order dependency in the relational model. *Theor Comput Sci* 26:146–195
- Han J, Kamber M (2006) *Data mining: concepts and techniques*, 2nd edn. Morgan Kaufman, San Diego
- Han J, Cai, Y, Cercone, N (1992) Knowledge discovery in databases: an attribute-oriented approach, *Proceedings of 18th VLDB Conference*, Vancouver, Brit. Columbia, pp 547–559

- Höller J, Tsiatsis V, Mulligan C, Karnouskos S, Avesand S, Boyle D (2014) From machine-to-machine to the internet of things: introduction to a new age of intelligence. Academic Press, Waltham
- Jing L, Wenwen Z (2016) Overview on the using rough set theory on GIS spatial relationships constraint. *Int J Adv Res Artif Intell*:11–15
- Klir GJ, Folger TA (1988) Fuzzy sets, uncertainty, and information. Prentice Hall, Englewood Cliffs
- Ligeza A (2002) Granular sets and granular relation. *Intelligent information systems*. Physica Verlag, Heidelberg, pp 331–340
- Lin TY (1997) Granular computing: from rough sets and neighborhood systems to information granulation and computing in words. *Eur Congr Intell Tech Soft Comput* 8-12:1602–1606
- Lin TY (1999) Granular computing: fuzzy logic and rough sets. In: Zadeh L, Kacprzyk J (eds) *Computing with words in information/intelligent systems*. Physica-Verlag, Heidelberg, pp 183–200
- Makinouchi A (1977) A consideration on normal form of not-necessarily normalized relation in the relational data model. In: *Proceedings of the 3rd international conference VLDB*, pp 447–453
- Mendel J (2017) *Uncertain rule-based fuzzy systems*, 2nd edn. Springer, Berlin
- Mendel J, John R (2002) Type-2 fuzzy sets made simple. *IEEE Trans Fuzzy Sets* 10:117–127
- Nanda S, Majumdar S (1992) Fuzzy rough sets. *Fuzzy Sets Syst* 45:157–160
- Ola A, Ozsoyoglu G (1993) Incomplete relational database models based on intervals. *IEEE Trans Knowl Data Eng* 5:293–308
- Parsons S (2001) *Qualitative methods for reasoning under uncertainty*. MIT Press, Cambridge
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11:341–356
- Pawlak Z (1984) Rough sets. *Int J Man-Mach Stud* 21:127–134
- Pawlak Z (1985) Rough sets and fuzzy sets. *Fuzzy Sets Syst* 17:99–102
- Pawlak Z (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Norwell
- Pedrycz W, Gomide F (1996) *An introduction to fuzzy sets: analysis and design*. MIT Press, Boston
- Petry F (1996) *Fuzzy databases: principles and applications*. Kluwer Press, Boston
- Petry F, Robinson V, Cobb M (2005) *Fuzzy modeling with spatial information for geographic problems*. Springer, Berlin/Heidelberg
- Petry F, Elmore P, Yager R (2015) Combining uncertain information of differing modalities. *Inf Sci* 322:237–256
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106
- Raschia G, Mouaddib N (2002) SAINTETIQ: a fuzzy set-based approach to database summarization. *Fuzzy Sets Syst* 129:137–162
- Roth M, Korth H, Batory D (1987) SQL/NF: a query language for non-1NF databases. *Inf Syst* 12:99–114
- Sent D, van de Gaag L (2007) In: Carbonell J, Siebarn J (eds) *On the behavior of information measures for test selection*. Lecture notes in AI 4594. Springer, Berlin
- Shafer G (1976) *A mathematical theory of evidence*. Princeton University Press, Princeton
- Shannon CL (1948) The mathematical theory of communication. *Bell Syst Tech J* 27:379–422
- Shi W, Wang S, Li D, Wang X (2003) Uncertainty-based spatial data mining. *Proceedings of Asia GIS Association*, Wuhan, pp 124–135
- Srinivasan P (1991) The importance of rough approximations for information retrieval. *Int J Man-Mach Stud* 34:657–671
- Stankovic J (2014) Research directions for the internet of things. *IEEE Internet Things J* 1(1):3–9
- Tavana M, Liu W, Elmore P, Petry F, Bourgeois BS (2016) A practical taxonomy of methods and literature for managing uncertain spatial data in geographic information systems. *Measurement* 82:123–162
- Wang S, Li D, Shi W, Wang X (2002) Rough spatial description, *International Archives of Photogrammetry and Remote Sensing*, XXXII, Commission II, pp 503–510
- Worboys M (1998a) Computation with imprecise geospatial data. *Comput Environ Urban Syst* 22:85–106
- Worboys M (1998b) Imprecision in finite resolution spatial data. *GeoInformatica* 2:257–280
- Wygralak M (1989) Rough sets and fuzzy sets—some remarks on interrelations. *Fuzzy Sets Syst* 29:241–243
- Yager R (1982) Measuring tranquility and anxiety in decision making. *Int J Gen Syst* 8:139–146
- Yager R (1992) On the specificity of a possibility distribution. *Fuzzy Sets Syst* 50:279–292
- Yager R (1995) Measures of entropy and fuzziness related to aggregation operators. *Inf Sci* 82:147–166
- Yager R (2012) Conditional approach to possibility-probability fusion. *IEEE Trans Fuzzy Syst* 20:46–56
- Yager R, Petry F (2016) An intelligent quality based approach to fusing multi-source probabilistic information. *Info Fusion* 31:127–136
- Zadeh L (1965) Fuzzy Sets. *Inf Control* 8:338–353
- Zadeh L (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28
- Zvieli A, Chen P (1986) Entity-relationship modeling and fuzzy databases. In: *Proceedings of international conference on data engineering*, pp 320–327

Books and Reviews

- Aczel J, Daroczy Z (1975) *On measures of information and their characterization*. Academic Press, New York
- Angryk R, Petry F (2007) Attribute-oriented fuzzy generalization in proximity and similarity-based relation database systems. *Int J Intell Syst* 22:763–781
- Arora G, Petry F, Beaubouef T (1997) Information measure of type β under similarity relations, sixth IEEE

- international conference on fuzzy systems Barcelona, pp 857–862
- Arora G, Petry F, Beaubouef T (2001) A note on new parametric measures of information for fuzzy sets. *J Combinatorics, Info Syst Sci* 26:167–174
- Beaubouef T, Petry F (2001a) Vague regions and spatial relationships: a rough set approach. In: Fourth international conference on computational intelligence and multimedia applications, Yokosuka City, pp 313–318
- Beaubouef T, Petry F (2001b) Vagueness in spatial data: rough set and egg-yolk approaches. In: 14th international conference on industrial & engineering applications of artificial intelligence, pp 367–373
- Beaubouef T, Petry F (2003) In: Bouchon-Meunier B, Foulloy L, Yager R (eds) *Rough set uncertainty in an object oriented data model, intelligent Systems for Information Processing: from representation to applications*. Elsevier, Amsterdam, pp 37–46
- Beaubouef T, Petry F (2005a) Normalization in a rough relational database, international conference on rough sets, fuzzy sets, data mining and granular computing, pp 257–265
- Beaubouef T, Petry F (2005b) Representation of spatial data in an OODB using rough and fuzzy set modeling. *Soft Comput J* 9:364–373
- Beaubouef T, Petry F (2007) An attribute-oriented approach for knowledge discovery in rough relational databases, *proc FLAIRS'07*, pp 507–508
- Beaubouef T, Petry F, Arora G (1998) Information measures for rough and fuzzy sets and application to uncertainty in relational databases. In: Pal S, Skowron A (eds) *Rough-fuzzy hybridization: a new trend in decision-making*. Springer, Singapore, pp 200–214
- Beaubouef T, Ladner R, Petry F (2004) Rough set spatial data modeling for data mining. *Int J Intell Syst* 19:567–584
- Beaubouef T, Petry F, Ladner R (2007) Spatial data methods and vague regions: a rough set approach. *Appl Soft Comput J* 7:425–440
- Buckles B, Petry F (1982) Security and fuzzy databases. In: *Proceedings 1982 IEEE international conference on cybernetics and society*, pp 622–625
- Codd E (1970) A relational model of data for large shared data banks. *Commun ACM* 13:377–387
- Ebanks B (1983) On measures of fuzziness and their representations. *J Math Anal Appl* 94:24–37
- Grzymala-Busse J (1991) *Managing uncertainty in expert systems*. Kluwer Academic Publishers, Boston
- Han J, Nishio S, Kawano H, Wang W (1998) Generalization-based data Mining in Object-Oriented Databases Using an object-cube model. *Data Knowl Eng* 25:55–97
- Havrda J, Charvat F (1967) Quantification methods of classification processes: concepts of structural α entropy. *Kybernetika* 3:149–172
- Kapur J, Kesavan H (1992) *Entropy optimization principles with applications*. Academic Press, New York
- Slowinski R (1992) A generalization of the indiscernibility relation for rough sets analysis of quantitative information. In: 1st international workshop on rough sets: state of the art and perspectives, Poland. In: pp 41–48

Rough Set Data Analysis

Shusaku Tsumoto

Department of Medical Informatics, Faculty of
Medicine, Shimane University, Enya-cho Izumo
City, Shimane, Japan

Article Outline

Introduction
Focusing Mechanism
Definition of Rules
Algorithms for Rule Induction
Experimental Results
What Is Discovered?
Rule Discovery as Knowledge Acquisition and
Decision Support
Discussion
Conclusions
Bibliography

Introduction

Rule-induction methods are classified into two categories, induction of deterministic rules and of probabilistic ones (Michalski et al. 1986; Pawlak 1991; Quinlan 1993; Tsumoto and Tanaka 1996).

On one hand, deterministic rules are described as *if-then* rules, which can be viewed as propositions. From the set-theoretical point of view, a set of examples supporting the conditional part of a deterministic rule, denoted by C , is a subset of a set whose examples belong to the consequence part, denoted by D . That is, the relation $C \subseteq D$ holds and deterministic rules are supported only by positive examples in a data set.

On the other hand, probabilistic rules are *if-then* rules with probabilistic information (Tsumoto and Tanaka 1996).

When a classical proposition will not hold for C and D , C is not a subset of D but closely

overlapped with D . That is, the relations $C \cap D \neq \emptyset$ and $|C \cap D| / |C| \geq \delta$ will hold in this case, where the threshold δ is the degree of closeness of overlapping sets, which will be given by domain experts. (For more information, see section “[Definition of Rules](#)”).

Thus, probabilistic rules are supported by a large number of positive examples and a few negative examples.

The common feature of both deterministic and probabilistic rules is that they deduce their consequence positively if an example satisfies their conditional parts. We call the reasoning by these rules *positive reasoning*.

However, medical experts use not only positive reasoning but also *negative reasoning* for selection of candidates, which is represented as *if-then* rules whose consequences include negative terms.

For example, when a patient who complains of headache does not have a throbbing pain, migraine should not be suspected with a high probability. Thus, negative reasoning also plays an important role in cutting the search space of a differential diagnosis process (Tsumoto and Tanaka 1996). Thus, medical reasoning includes both positive and negative reasoning, though conventional rule-induction methods do not reflect this aspect. This is one of the reasons medical experts have difficulty in interpreting induced rules, and interpreting rules for a discovery procedure does not proceed easily. Therefore, negative rules should be induced from databases in order not only to induce rules reflecting experts’ decision processes, but also to induce rules that will be easier for domain experts to interpret, both of which are important to enhance the discovery process done by the cooperation of medical experts and computers.

In this chapter, first the characteristics of medical reasoning are discussed and then two kinds of rules, positive rules and negative rules, are introduced as a model of medical reasoning. Interestingly, from the set-theoretic point of view, sets of

examples supporting both rules correspond to the lower and upper approximations in rough sets (Pawlak 1991). On the other hand, from the viewpoint of propositional logic, both positive and negative rules are defined as classical propositions or deterministic rules with two probabilistic measures, classification accuracy, and coverage.

Second, two algorithms for induction of positive and negative rules are introduced, defined as search procedures using accuracy and coverage as evaluation indices.

Finally, the proposed method is evaluated on several medical databases. The experimental results show that the induced rules correctly represent experts' knowledge. In addition, several interesting patterns are discovered.

Focusing Mechanism

One of the characteristics in medical reasoning is a focusing mechanism, which is used to select the final diagnosis from many candidates (Tsumoto and Tanaka 1996; Tsumoto 1998a). For example, in differential diagnosis of headache, more than 60 diseases will be checked by present history, physical examinations, and laboratory examinations. In diagnostic procedures, a candidate is excluded if a symptom necessary to diagnose is not observed.

This style of reasoning consists of the following two processes: exclusive reasoning and inclusive reasoning. Relations of this diagnostic model with another diagnostic model are discussed in (Tsumoto 1998b). The diagnostic procedure

proceeds as follows (Fig. 1): First, exclusive reasoning excludes a disease from candidates when a patient does not have a symptom that is necessary to diagnose that disease.

Second, inclusive reasoning suspects a disease in the output of the exclusive process when a patient has symptoms specific to a disease.

These two steps are modeled as two kinds of rules, negative rules (or exclusive rules) and positive rules; the former corresponds to exclusive reasoning, the latter to inclusive reasoning.

In the next two sections, these two rules are represented as special kinds of probabilistic rules.

Definition of Rules

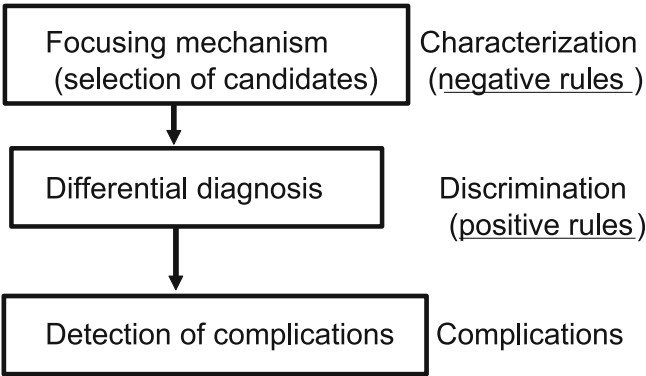
Rough Sets

In the following sections, we use the following notation introduced by Grzymala-Busse and Skowron (Skowron and Grzymala-Busse 1994), based on rough set theory (Pawlak 1991).

These notations are illustrated by a small data set shown in Table 1, which includes symptoms exhibited by six patients who complained of headache.

Let U denote a nonempty finite set called the universe and A denote a nonempty, finite set of attributes, i. e., $a : U \rightarrow V_a$ for $a \in A$, where V_a is called the domain of a , respectively. Then a decision table is defined as an information system, $A = (U, A \cup \{d\})$. For example, Table 1 is an information system with $U = \{1, 2, 3, 4, 5, 6\}$ and $A =$

Rough Set Data Analysis,
Fig. 1 Illustration of
focusing mechanism



Rough Set Data Analysis, Table 1 An example of a data set

No.	Age	Location	Nature	Prodrome	Nausea	M1	Class
1	50–59	ocular	persistent	no	no	yes	m.c.h.
2	40–49	whole	persistent	no	no	yes	m.c.h.
3	40–49	lateral	throbbing	no	yes	no	migra
4	40–49	whole	throbbing	yes	yes	no	migra
5	40–49	whole	radiating	no	no	yes	m.c.h.
6	50–59	whole	persistent	no	yes	yes	psycho

M1 tenderness of M1, m.c.h. muscle contraction headache, migra migraine, psycho psychological pain

$\{\text{age, location, nature, prodrome, nausea, } M1\}$ and $d = \text{class}$. For $\text{location} \in A$, V_{location} is defined as $\{\text{ocular, lateral, whole}\}$.

The atomic formulas over $B \subseteq A \cup \{d\}$ and V are expressions of the form $[a = v]$, called descriptors over $\mathbf{B}$, where $a \in B$ and $v \in V_a$. The set $F(B, V)$ of formulas over $\mathbf{B}$ is the least set containing all atomic formulas over B and closed with respect to disjunction, conjunction, and negation. For example, $[\text{location} = \text{ocular}]$ is a descriptor of B .

For each $f \in F(B, V)$, f_A denotes the meaning of f in A , i. e., the set of all objects in U with property f , defined inductively as follows:

1. If f is of the form $[a = v]$, then $f_A = \{s \in U \mid a(s) = v\}$.
2. $(f \wedge g)_A = f_A \cap g_A$; $(f \vee g)_A = f_A \cup g_A$; $(\neg f)_A = U - f_A$.

For example, $f = [\text{location} = \text{whole}]$ and $f_A = \{2, 4, 5, 6\}$. As an example of a conjunctive formula, $g = [\text{location} = \text{whole}] \wedge [\text{nausea} = \text{no}]$ is a descriptor of U and f_A is equal to $g_{\text{location,nausea}} = \{2, 5\}$.

Classification Accuracy and Coverage

Definition of Accuracy and Coverage By use of the preceding framework, classification accuracy and coverage, or true positive rate are defined as follows.

Definition 1 Let R and D denote a formula in $F(B, V)$ and a set of objects that belong to a

decision d . Classification accuracy and coverage (true positive rate) for $R \rightarrow d$ is defined as:

$$\alpha_R(D) = \frac{|R_A \cap D|}{|R_A|} (= P(D|R)) \text{ and}$$

$$\kappa_R(D) = \frac{|R_A \cap D|}{|D|} (= P(R|D)),$$

where $|S|$, $\alpha_R(D)$, $\kappa_R(D)$, and $P(S)$ denote the cardinality of a set S , a classification accuracy of R as to classification of D , and coverage (a true positive rate of R to D), and probability of S , respectively.

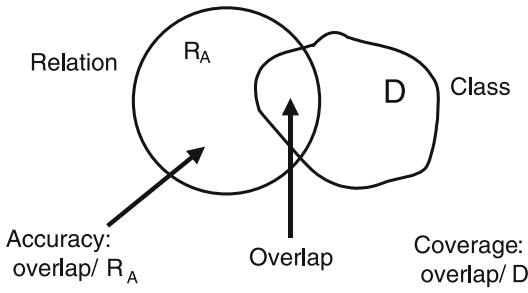
Figure 2 depicts the Venn diagram of relations between accuracy and coverage. Accuracy views the overlapped region $|R_A \cap D|$ from the meaning of a relation R . On the other hand, coverage views the overlapped region from the meaning of a concept D .

In the preceding example, when R and D are set to $[\text{nau} = \text{yes}]$ and $[\text{class} = \text{migraine}]$, $\alpha_R(D) = 2/3 = 0.67$ and $\kappa_R(D) = 2/2 = 1.0$.

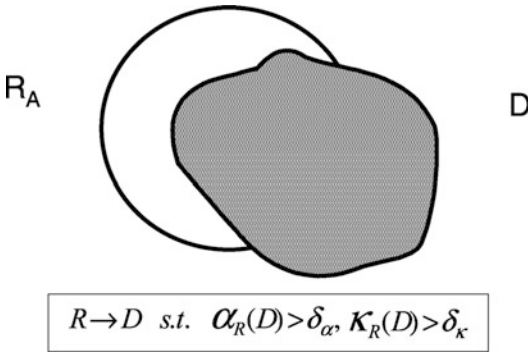
It is notable that $\alpha_R(D)$ measures the degree of the sufficiency of a proposition, $R \rightarrow D$, and that $\kappa_R(D)$ measures the degree of its necessity. For example, if $\alpha_R(D)$ is equal to 1.0, then $R \rightarrow D$ is true. On the other hand, if $\kappa_R(D)$ is equal to 1.0, then $D \rightarrow R$ is true. Thus, if both measures are 1.0, then $R \leftrightarrow D$.

Probabilistic Rules

By use of accuracy and coverage, a probabilistic rule is defined as:



Rough Set Data Analysis, Fig. 2 Venn diagram of accuracy and coverage



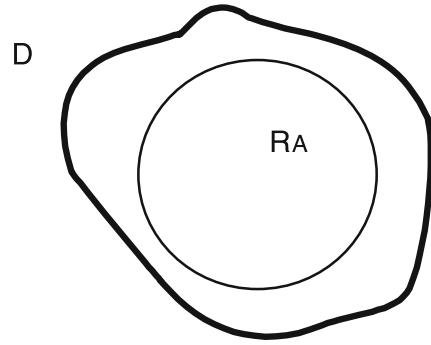
Rough Set Data Analysis, Fig. 3 Venn diagram for probabilistic rules

$$R \xrightarrow{\alpha, \kappa} d \quad \text{s.t.} \quad R = \bigwedge_j [a_j = v_k], \quad \alpha_R(D) \geq \delta_\alpha \\ \text{and} \quad \kappa_R(D) \geq \delta_\kappa.$$

If the thresholds for accuracy and coverage are set to high values, the meaning of the conditional part of probabilistic rules corresponds to the highly overlapped region. Figure 3 depicts the Venn diagram of probabilistic rules with highly overlapped region.

This rule is a kind of probabilistic proposition with two statistical measures, which is an extension of Ziarko's variable precision model (VPRS) (Ziarko 1993).¹

It is also notable that both a positive rule and a negative rule are defined as special cases of this rule, as shown in the next sections.



Rough Set Data Analysis, Fig. 4 Venn diagram of positive rules

Positive Rules

A positive rule is defined as a rule supported by only positive examples, the classification accuracy of which is equal to 1.0.

It is notable that the set supporting this rule corresponds to a subset of the lower approximation of a target concept, which is introduced in rough sets (Pawlak 1991). Thus, a positive rule is represented as:

$$R \rightarrow d \quad \text{s.t.} \quad R = \bigwedge_j [a_j = v_k], \quad \alpha_R(D) = 1.0.$$

Figure 4 shows the Venn diagram of a positive rule. As shown in this figure, the meaning of R is a subset of that of D . This diagram is exactly equivalent to the classic proposition $R \rightarrow d$.

In the preceding example, one positive rule of m.c.h. (muscle contraction headache) is:

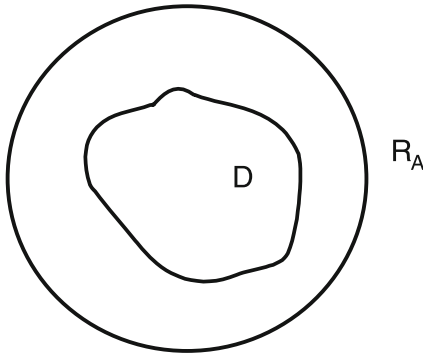
$$[\text{nausea} = \text{no}] \rightarrow \text{m.c.h.} \quad \alpha = 3/3 = 1.0.$$

This positive rule is often called a deterministic rule. However, we use the term, positive (deterministic) rules, because a deterministic rule supported only by negative examples, called a negative rule, is introduced in the next section.

Negative Rules

Before defining a negative rule, let us first introduce an exclusive rule, the contrapositive of a negative rule (Tsumoto and Tanaka 1996). An exclusive rule is defined as a rule supported by all the positive examples, the coverage of which is

¹This probabilistic rule is also a kind of *rough modus ponens* (Pawlak 1998)



Rough Set Data Analysis, Fig. 5 Venn diagram of exclusive rules

equal to 1.0. That is, an exclusive rule represents the necessity condition of a decision.

It is notable that the set supporting an exclusive rule corresponds to the upper approximation of a target concept, which is introduced in rough sets (Pawlak 1991). Thus, an exclusive rule is represented as:

$$R \rightarrow d \quad \text{s.t.} \quad R = \bigvee_j [a_j = v_k], \quad \kappa_R(D) = 1.0.$$

Figure 5 shows the Venn diagram of an exclusive rule. As shown in this figure, the meaning of R is a superset of that of D . This diagram is exactly equivalent to the classic proposition $d \rightarrow R$.

In the preceding example, the exclusive rule of m.c.h. is:

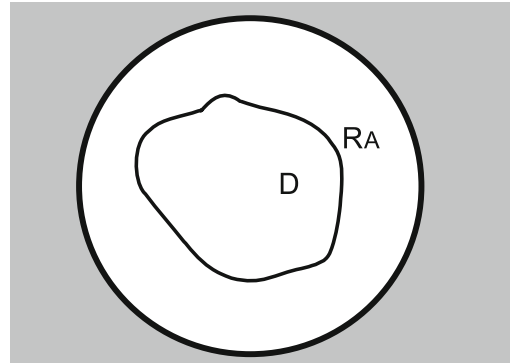
$$[M1 = \text{yes}] \vee [\text{nau} = \text{no}] \rightarrow \text{m.c.h.} \quad \kappa = 1.0,$$

From the viewpoint of propositional logic, an exclusive rule should be represented as:

$$d \rightarrow \bigvee_j [a_j = v_k],$$

because the condition of an exclusive rule corresponds to the necessity condition of conclusion d . Thus, it is easy to see that a negative rule is defined as the contrapositive of an exclusive rule:

$$\bigwedge_j \neg [a_j = v_k] \rightarrow \neg d,$$



Rough Set Data Analysis, Fig. 6 Venn diagram of negative rules

which means that if a case does not satisfy any attribute value pairs in the condition of a negative rule, then we can exclude a decision d from candidates.

For example, the negative rule of m.c.h. is:

$$\neg[M1 = \text{yes}] \wedge \neg[\text{nausea} = \text{no}] \rightarrow \neg \text{m.c.h.}$$

In summary, a negative rule is defined as:

$$\begin{aligned} \bigwedge_j \neg [a_j = v_k] &\rightarrow \neg d \quad \text{s.t.} \\ \forall [a_j = v_k] &\kappa_{[a_j = v_k]}(D) = 1.0, \end{aligned}$$

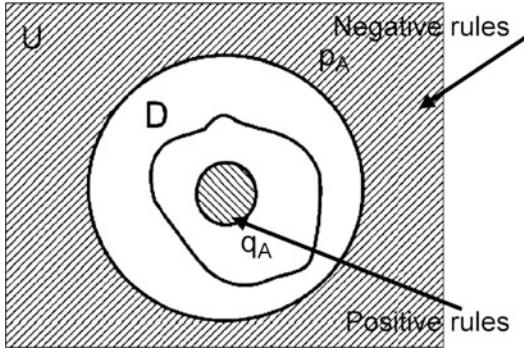
where D denotes a set of samples that belong to a class d . Figure 6 shows the Venn diagram of a negative rule. As shown in this figure, it is notable that this negative region is the “positive region” of “negative concept.”

Negative rules should also be included in a category of deterministic rules, because their coverage, a measure of negative concepts, is equal to 1.0. It is also notable that the set supporting a negative rule corresponds to a subset of negative region, which is introduced in rough sets (Pawlak 1991).

In summary, positive and negative rules correspond to positive and negative regions defined in rough sets. Figure 7 shows the Venn diagram of those rules.

Algorithms for Rule Induction

The contrapositive of a negative rule, an exclusive rule, is induced as an exclusive rule by the modification of the algorithm introduced in



Rough Set Data Analysis, Fig. 7 Venn diagram of defined rules

PRIMEROSE-REX (Tsumoto and Tanaka 1996), as shown in Fig. 8. This algorithm works as follows.

1. First it selects a descriptor $[a_i = v_j]$ from the list of attribute-value pairs, denoted by L .
2. Then it checks whether this descriptor overlaps with a set of positive examples, denoted by D .
3. If so, this descriptor is included in a list of candidates for positive rules and the algorithm checks whether its coverage is equal to 1.0. If the coverage is equal to 1.0, then this descriptor is added to R_{er} , the formula for the conditional part of the exclusive rule of D .
4. Then $[a_i = v_j]$ is deleted from the list L . This procedure, from (1) to (4), will continue unless L is empty.
5. Finally, when L is empty, this algorithm generates negative rules by taking the contrapositive of induced exclusive rules.

procedure Exclusive and Negative Rules;

var

L : List;

/* A list of elementary attribute-value pairs */

begin

$L := P_0$;

/* P_0 : A list of elementary attribute-value pairs given in a database */

while ($L \neq \{\}$) **do**

begin

Select one pair $[a_i = v_j]$ from L ;

if ($[a_i = v_j]_A \cap D \neq \phi$) **then do** /* D : positive examples of a target class d */

begin

$L_{ir} := L_{ir} + [a_i = v_j]$; /* Candidates for Positive Rules */

if ($\kappa_{[a_i = v_j]}(D) = 1.0$)

then $R_{er} := R_{er} \wedge [a_i = v_j]$;

/* Include $[a_i = v_j]$ into the formula of Exclusive Rule */

end

$L := L - [a_i = v_j]$;

end

Construct Negative Rules:

Take the contrapositive of R_{er} .

end {Exclusive and Negative Rules};

Rough Set Data Analysis, Fig. 8 Induction of exclusive and negative rules

```

procedure Positive Rules;
  var
     $i$  : integer;   $M, L_i$  : List;
  begin
     $L_1 := L_{ir}$ ;
    /*  $L_{ir}$ : A list of candidates generated by induction of exclusive rules */
     $i := 1$ ;   $M := \{\}$ ;
    for  $i := 1$  to  $n$  do
      /*  $n$ : Total number of attributes given
        in a database */
      begin
        while ( $L_i \neq \{\}$ ) do
          begin
            Select one pair  $R = \wedge[a_i = v_j]$  from  $L_i$ ;
             $L_i := L_i - \{R\}$ ;
            if ( $\alpha_R(D) > \delta_\alpha$ )
              then do  $S_{ir} := S_{ir} + \{R\}$ ;
            /* Include  $R$  in a list of the Positive Rules */
            else  $M := M + \{R\}$ ;
          end
           $L_{i+1} :=$  (A list of the whole combination of the conjunction formulae in  $M$ );
        end
      end {Positive Rules};

```

Rough Set Data Analysis, Fig. 9 Induction of positive rules

On the other hand, positive rules are induced as inclusive rules by the algorithm introduced in PRIMEROSE-REX (Tsumoto and Tanaka 1996), as shown in Fig. 9. For induction of positive rules, the threshold of accuracy and coverage is set to 1.0 and 0.0, respectively.

This algorithm works in the following way.

1. First it substitutes L_1 , which denotes a list of formulas composed of only one descriptor, with the list L_{er} generated by the former algorithm shown in Fig. 9.
2. Then until L_1 becomes empty, the following steps will continue: (a) A formula $[a_i = v_j]$ is removed from L_1 . (b) Then the algorithm checks whether $\alpha_R(D)$ is larger than the threshold. (For induction of positive rules, this is equal to checking whether $\alpha_R(D)$ is equal to

1.0.) If so, then this formula is included a list of the conditional parts of positive rules. Otherwise, it will be included in M , which is used for making conjunctions.

3. When L_1 is empty, the next list L_2 is generated from the list M .

Experimental Results

For experimental evaluation, a new system, called PRIMEROSE-REX2 (Probabilistic Rule Induction Method for Rules of Expert System ver. 2.0), was developed, where the algorithms discussed in section “[Algorithms for Rule Induction](#)” were implemented.

PRIMEROSE-REX2 was applied to the following three medical domains:

- 1. Headache (RHINOS domain), whose training samples consist of 52,119 samples, 45 classes, and 147 attributes;
- 2. Cerebrovascular diseases (CVD), whose training samples consist of 7620 samples, 22 classes, and 285 attributes; and
- 3. Meningitis, whose training samples consist of 1211 samples, 4 classes, and 41 attributes (Table 2).

For evaluation, we used the following two types of experiments. One experiment was to evaluate the predictive accuracy using the cross-validation method, which is often used in the machine-learning literature (Shavlik and Dietterich 1990). The other experiment was to evaluate the induced rules by medical experts and to check whether these rules lead to a new discovery.

Performance of Rules Obtained

For comparison of performance, the experiments are conducted by the following four procedures. First, rules are acquired manually from experts. Second, the data sets are randomly split into new training samples and new test samples. Third, PRIMROSE-REX2, conventional rule-induction methods, AQ15 (Michalski et al. 1986) and C4.5 (Quinlan 1993) are applied to the new training samples for rule generation. Fourth, the induced rules and rules acquired

from experts are tested using new test samples. The second through fourth steps are repeated 100 times, and the average classification accuracy over 100 trials is computed. This process is a variant of repeated two-fold cross-validation, introduced in (Tsumoto and Tanaka 1996).

Experimental results (performance) are shown in Table 3. The first and second rows show the results obtained using PRIMROSE-REX2; the results in the first row are derived using both positive and negative rules and those in the second row are derived by only positive rules. The third row shows the results derived from medical experts. For comparison, we compare the classification accuracy of C4.5 and AQ-15, which is shown in the fourth and fifth rows.

These results show that the combination of positive and negative rules outperforms positive rules, although it is a little worse than medical experts' rules.

What Is Discovered?

Interesting Rules Are Very Few

One of the most important observations is that there were very few rules interesting or unexpected to medical experts compared to the number of rules extracted from the data sets.

Table 4 shows the mentioned results. The second column denotes the number of positive and negative rules obtained from each data set. The third column denotes the number of positive and negative rules interesting or unexpected for domain experts. For example, the first row shows that the number of induced positive rules in headache is 24,335, but the number of interesting rules are 114, which shows that only 0.47% of rules are interesting for domain experts.

Rough Set Data Analysis, Table 2 Databases

Domain	Samples	Classes	Attributes
Headache	52,119	45	147
CVD	7620	22	285
Meningitis	1211	4	41

Rough Set Data Analysis, Table 3 Experimental results (accuracy: averaged)

Method	Headache	CVD	Meningitis
PRIMROSE-REX2 (positive+negative)	91.3%	89.3%	92.5%
PRIMROSE-REX2 (positive)	68.3%	71.3%	74.5%
Experts	95.0%	92.9%	93.2%
PR-REX	88.3%	84.3%	82.5%
C4.5	85.8%	79.7%	81.4%
AQ15	86.2%	78.9%	82.5%

Rough Set Data Analysis, Table 4 Number of extracted rules and interesting rules

	Induced positive rules	Interesting rules
Headache	24,335	114 (0.47%)
CVD	14,715	106 (0.72%)
Meningitis	1922	40 (2%)
	Induced negative rules	Interesting rules
Headache	12,113	120 (0.99%)
CVD	7231	155 (2.1%)
Meningitis	77	5 (6.5%)

This table shows that the number of interesting rules are very few: even in the case of meningitis, only 6.5% are interesting, which suggests that the interpretation part of domain experts is very hard for the knowledge discovery process.

Next, we show several examples of rules that are interesting to domain experts.

Positive Rules in Differential Diagnosis of Headache

In the domain of differential diagnosis of headache, the following interesting positive rules were found.

$$[Age < 20] \wedge [History : paroxysmal] \rightarrow \text{Common Migraine}(\text{Coverage} : 0.75)$$

This rule is said to be interesting if it is compared with the following rule:

$$[Age > 40] \wedge [History : paroxysmal] \rightarrow \text{Classic Migraine}(\text{Coverage} : 0.72)$$

These two rules include two parts; although the values of the attribute “Age” are different, those of the attribute “History” are the same. This suggests that the attribute “Age” is important for differential diagnosis between common migraine and classic migraine.

Negative Rules in Differential Diagnosis of Headache

In the domain of differential diagnosis of headache, the following interesting negative rule was found.

$$\neg[Nature : Persistent] \wedge \neg[History : acute] \wedge \neg[History : paroxysmal] \wedge \neg[Neck Stiffness : yes] \rightarrow \neg\text{Common Migraine}$$

It is notable that it is difficult even for domain experts to interpret the interestingness of this rule if only this rule is shown. The domain experts pointed out that the rule is interesting when compared with the following rules.

$$[Nature : Persistent] \wedge [History : acute] \wedge [Neck Stiffness : yes] \rightarrow \text{Meningitis}[Nature : Persistent] \wedge [History : chronic] \rightarrow \text{Brain Tumor}$$

This means that the rule includes information that is very important for differential diagnosis between common migraine and meningitis or brain tumor. Note that these two rules are very similar to the negative rule except for the negative symbols.

Positive Rules in Meningitis

In the domain of meningitis, the following positive rules, which medical experts do not expect, are obtained.

$$[WBC < 12000] \wedge [Sex = Female] \wedge [Age < 40] \wedge [CSF_CELL < 1000] \rightarrow \text{Virus}(\text{Coverage} : 0.91) \\ [Age \geq 40] \wedge [WBC \geq 8000] \wedge [Sex = Male] \wedge [CSF_CELL \geq 1000] \rightarrow \text{Bacteria}(\text{Coverage} : 0.64)$$

The former rule means that if WBC (white blood cell count) is less than 12,000, the gender of a patient is female, the age is less than 40, and CSF_CELL (cell count of cerebrospinal fluid), then the type of meningitis is Viral. The latter means that the age of a patient is less than 40, WBC is larger than 8000, the gender is male, and CSF_CELL is larger than 1000, then the type of meningitis is Bacterial.

The most interesting points are that these rules included information about age and gender, which often seems to be unimportant attributes for differential diagnosis of meningitis. The first discovery was that women did not often suffer from bacterial infection compared with men, because such relationships between gender and meningitis has not been discussed in medical context (Adams and Victor 1993). By a close examination of the

database on meningitis, it was found that most of the patients suffered from chronic diseases, such as DM, LC, and sinusitis, which are the risk factors of bacterial meningitis. The second discovery was that $[age < 40]$ was also an important factor not to suspect viral meningitis, which also matches the fact that most old people suffer from chronic diseases. These results were also reevaluated in medical practice. Recently, the preceding two rules were checked by an additional 21 cases who suffered from meningitis (15 cases viral, 6 cases bacterial meningitis.) Surprisingly, the rules misclassified only three cases (two viral, the other bacterial), that is, the total accuracy was equal to $18/21 = 85.7\%$, and the accuracies for viral and bacterial meningitis were equal to $13/15 = 86.7\%$ and $5/6 = 83.3\%$. The reasons for misclassification are the following: a case of bacterial infection involved a patient who had a severe immunodeficiency, although he is very young. Two cases of viral infection involved patients who suffered from herpes zoster. It is notable that even those misclassified cases could be explained from the viewpoint of the immunodeficiency: that is, it was confirmed that immunodeficiency is a key factor for meningitis.

The validation of these rules is still ongoing; it will be reported in the near future.

Positive and Negative Rules in CVD

Concerning the database on CVD, several interesting rules were derived. The most interesting results were the following positive and negative rules for thalamus hemorrhage:

$$\begin{aligned} & [Sex = Female] \wedge [Hemiparesis = Left] \\ & \wedge [LOC : positive] \longrightarrow Thalamus \\ & \neg[Risk : Hypertension] \wedge \neg[Sensory = no] \\ & \longrightarrow \neg Thalamus \end{aligned}$$

The former rule means that if the gender of a patient is female and he or she suffered from the left hemi-paresis ($[Hemiparesis = Left]$) and loss of consciousness ($[LOC:positive]$), then the focus of CVD is thalamus. The latter rule means that if

he or she suffers neither from hypertension ($[Risk: Hypertension]$) or sensory disturbance ($[Sensory = no]$), then the focus of CVD is thalamus.

Interestingly, LOC (loss of consciousness) under the condition of $[Gender = Female] \wedge [Hemiparesis = Left]$ was found to be an important factor to diagnose thalamic damage. In this domain, any strong correlations between these attributes and others, like the database of meningitis, have not been found yet. It will be our future work to find what factor is behind these rules.

Rule Discovery as Knowledge Acquisition and Decision Support

Expert System: RH

Another point of discovery of rules is automated knowledge acquisition from databases. Knowledge acquisition is referred to as a bottleneck problem in development of expert systems (Buchnan and Shortliffe 1984), which has not fully been solved and is expected to be solved by induction of rules from databases. However, there are few papers that discuss the evaluation of discovered rules from the viewpoint of knowledge acquisition (Tsumoto 1998b).

For this purpose, we have developed an expert system, called RH (rule-based system for headache) using the acquired knowledge. The reason for selecting the domain of headache is that earlier we developed an expert system RHINOS (rule-based headache information organizing system), which makes a differential diagnosis in headache (Matsumura et al. 1988). In this system, it takes about six months to acquire knowledge from domain experts. RH consists of two parts. First, it requires inputs and applies exclusive and negative rules to select candidates (focusing mechanism). Then, it requires additional inputs and applies positive rules for differential diagnosis between selected candidates. Finally, RH outputs diagnostic conclusions.

Rough Set Data Analysis, Table 5 Evaluation of RH (accuracy: averaged)

Method	Accuracy
PRIMEROSE-REX2 (positive and negative)	91.4% (851/930)
PRIMEROSE-REX2 (positive)	78.5% (729/930)
RHINOS (positive and negative)	93.5% (864/930)
RHINOS (positive)	82.8% (765/930)

Evaluation of RH

RH was evaluated in clinical practice with respect to its classification accuracy by using 930 patients who came to the outpatient clinic after the development of this system. Experimental results about classification accuracy are shown in Table 5. The first and second rows show the performance of rules obtained using PRIMROSE-REX2; the results in the first row are derived using both positive and negative rules and those in the second row are derived using only positive rules. The third and fourth rows show the results derived using both positive and negative rules and those by positive rules acquired directly from medical experts. These results show that the combination of positive and negative rules outperforms positive rules and gains almost the same performance as those by experts.

Discussion

Hierarchical Rules for Decision Support

One of the problems with rule induction is that conventional rule-induction methods cannot extract rules that plausibly represent experts' decision processes (Tsumoto 1998b). The description length of induced rules is too short, compared to the experts' rules. (It may be observed that this length part does not contribute much to the classification performance.)

For example, rule-induction methods introduced in this chapter induced the following common rule for muscle contraction headache from databases on differential diagnosis of headache:

[location = whole]
 $\wedge$ [Jolt Headache = no]
 $\wedge$ [Tenderness of M1 = yes]
 $\rightarrow$ muscle contraction headache

This rule is shorter than the following rule given by medical experts:

[Jolt Headache = no]
 $\wedge$ ([Tenderness of M0 = yes]
 $\vee$ [Tenderness of M1 = yes]
 $\vee$ [Tenderness of M2 = yes])
 $\wedge$ [Tenderness of B1 = no]
 $\wedge$ [Tenderness of B2 = no]
 $\wedge$ [Tenderness of B3 = no]
 $\wedge$ [Tenderness of C1 = no]
 $\wedge$ [Tenderness of C2 = no]
 $\wedge$ [Tenderness of C3 = no]
 $\wedge$ [Tenderness of C4 = no]
 $\rightarrow$ muscle contraction headache

These results suggest that conventional rule-induction methods do not reflect a mechanism of knowledge acquisition of medical experts.

Typically, rules acquired from medical experts are much longer than those induced from databases, the decision attributes of which are given by the same experts. This is because rule induction methods generally search for shorter rules, compared with decision tree induction. In the case of decision tree induction, the induced trees are sometimes too deep, and in order for the trees to be useful for learning, pruning and examination by experts are required. One of the main reasons rules are short and decision trees are sometimes long is that these patterns are generated by only one criteria, such as high accuracy or high information gain. The comparative study in this section suggests that experts should acquire rules by usage of several measures. Those characteristics of medical experts' rules are fully examined not by comparing between those rules for the same class, but by comparing experts' rules with those for another class. For example, a

classification rule for muscle contraction headache is given by:

$$\begin{aligned}
 &[\text{Jolt Headache} = \text{no}] \\
 &\wedge([\text{Tenderness of M0} = \text{yes}] \\
 &\vee[\text{Tenderness of M1} = \text{yes}] \\
 &\vee[\text{Tenderness of M2} = \text{yes}]) \\
 &\wedge[\text{Tenderness of B1} = \text{no}] \\
 &\wedge[\text{Tenderness of B2} = \text{no}] \\
 &\wedge[\text{Tenderness of B3} = \text{no}] \\
 &\wedge[\text{Tenderness of C1} = \text{no}] \\
 &\wedge[\text{Tenderness of C2} = \text{no}] \\
 &\wedge[\text{Tenderness of C3} = \text{no}] \\
 &\wedge[\text{Tenderness of C4} = \text{no}] \\
 &\rightarrow \text{muscle contraction headache}
 \end{aligned}$$

This rule is very similar to the following classification rule for disease of cervical spine:

$$\begin{aligned}
 &[\text{Jolt Headache} = \text{no}] \\
 &\wedge([\text{Tenderness of M0} = \text{yes}] \\
 &\vee[\text{Tenderness of M1} = \text{yes}] \\
 &\vee[\text{Tenderness of M2} = \text{yes}]) \\
 &\wedge([\text{Tenderness of B1} = \text{yes}] \\
 &\vee[\text{Tenderness of B2} = \text{yes}] \\
 &\vee[\text{Tenderness of B3} = \text{yes}] \\
 &\vee[\text{Tenderness of C1} = \text{yes}] \\
 &\vee[\text{Tenderness of C2} = \text{yes}] \\
 &\vee[\text{Tenderness of C3} = \text{yes}] \\
 &\vee[\text{Tenderness of C4} = \text{yes}]) \\
 &\rightarrow \text{disease of cervical spine}
 \end{aligned}$$

The differences between these two rules are attribute-value pairs, from tenderness of B1 to C4. Thus, these two rules can be simplified into the following form:

$$\begin{aligned}
 a_1 \wedge A_2 \wedge \neg A_3 &\rightarrow \text{musclecontractionheadache}, \\
 a_1 \wedge A_2 \wedge A_3 &\rightarrow \text{diseaseofcervicalspine}.
 \end{aligned}$$

The first two terms and the third one represent different reasoning. The first and second terms a_1 and A_2 are used to differentiate muscle contraction headache and disease of cervical spine from other

diseases. The third term A_3 is used to make a differential diagnosis between these two diseases. Thus, medical experts first select several diagnostic candidates, which are similar to each other, from many diseases and then make a final diagnosis from those candidates.

This problem has been partially solved; Tsumoto introduced a new approach for inducing these rules in (Tsumoto 1998c), as induction of hierarchical decision rules. In that paper, the characteristics of experts' rules are closely examined and a new approach to extract plausible rules is introduced, which consists of the following three procedures. First, the characterization of decision attributes (given classes) is done from databases and the classes are classified into several groups with respect to the characterization. Then two kinds of subrules, characterization rules for each group and discrimination rules for each class in the group, are induced. Finally, those two parts are integrated into one rule for each decision attribute. The proposed method was evaluated on a medical database, the experimental results of which show that induced rules correctly represent experts' decision processes.

This observation also suggests that medical experts implicitly look at the relation between rules for different concepts. Future work should discover the relations between induced rules.

Relations Between Rules

In (Tsumoto 2001), Tsumoto focuses on the characteristics of medical reasoning(focusing mechanism) and introduces three kinds of rules, positive rules, exclusive rules and total covering rules, as a model of medical reasoning, which is an extended formalization of rules defined in (Tsumoto 1998b).

Interestingly, from the set-theoretic point of view, sets of examples supporting these rules correspond to the lower and upper approximations in rough sets.

Furthermore, from the viewpoint of propositional logic, both inclusive and exclusive rules are defined as classical propositions, or deterministic rules with two probabilistic measures, classification accuracy and coverage. Total covering rules have several interesting relations with inclusive

and exclusive rules, which reflects the characteristics of medical reasoning.

Originally, a total covering rule is defined as a set of symptoms that can be observed in at least one case of a target disease. That is, this rule is defined as a collection of attribute-value pairs whose accuracy is larger than 0:

$$R \rightarrow d \quad \text{s.t.} \quad R = \bigvee_j [a_j = v_k], \quad \alpha_R(D) > 0.$$

From the definition of accuracy and coverage, this formula can be transformed into:

$$R \rightarrow d \quad \text{s.t.} \quad R = \bigvee_j [a_j = v_k], \quad \kappa_R(D) > 0.$$

For each attribute, the attribute-value pairs form a partition of U . Thus, for each attribute, total covering rules include a covering of all the positive examples. According to this property, the preceding formula is redefined as:

$$R \rightarrow d \quad \text{s.t.} \quad R = \bigvee_j R(a_j), \\ R(a_j) = \bigvee_k [a_j = v_k] \quad \text{s.t.} \quad \kappa_R(D) = 1.0.$$

It is notable that this definition is an extension of exclusive rules and this total covering rule can be written as:

$$d \rightarrow \bigvee_j \bigvee_k [a_j = v_k] \quad \text{s.t.} \quad \kappa_{[a_j=v_k]}(D) = 1.0.$$

Let $S(R)$ denote a set of attribute-value pairs of rule R . For each class d , let $R_{\text{pos}}(d)$, $R_{\text{ex}}(d)$, and $R_{\text{tc}}(d)$ denote the positive exclusive rule and total covering rules, respectively. Then

$$S(R_{\text{ex}}(d)) \subseteq S(R_{\text{tc}}(d)),$$

because a total covering rule can be viewed as an upper approximation of exclusive rules. It is also notable that this relation will hold in the relation between the positive rule (inclusive rule) and total covering rule. That is,

$$S(R_{\text{pos}}(d)) \subseteq S(R_{\text{tc}}(d)).$$

Thus, the total covering rule can be viewed as an upper approximation of inclusive rules. This relation also holds when $R_{\text{pos}}(d)$ is replaced with a

probabilistic rule, which shows that total covering rules are the weakest form of diagnostic rules.

In this way, rules that reflect the diagnostic reasoning of medical experts have sophisticated background from the viewpoint of set theory. Especially, the rough set framework provides a good tool for modeling such focusing mechanisms.

Our future work will investigate the relation between these three types of rules from the viewpoint of rough set theory.

Conclusions

In this chapter, the characteristics of two measures, classification accuracy and coverage, are discussed, which show that both measures are dual and that accuracy and coverage are measures of both positive and negative rules, respectively. Then an algorithm for induction of positive and negative rules is introduced.

The proposed method is evaluated on medical databases. The experimental results have demonstrated that the induced rules are able to correctly represent experts' knowledge. We also demonstrated that the method can discover several interesting patterns.

Bibliography

- Adams RD, Victor M (1993) Principles of neurology, 5th edn. McGraw-Hill, New York
- Buchnan BG, Shortliffe EH (eds) (1984) Rule-based expert systems. Addison-Wesley
- Matsumura Y, Matsunaga T, Hata Y, Kimura M, Matsumura H (1988) Consultation system for diagnoses of headache and facial pain. Med Inform RHINOS 11:145–157
- Michalski RS, Mozetic I, Hong J, Lavrac N (1986) The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. In: Proceedings of the 5th national conference on artificial intelligence. AAAI Press, Palo Alto, pp 1041–1045
- Pawlak Z (1991) Rough sets. Kluwer, Dordrecht
- Pawlak Z (1998) Rough modus ponens. In: Proceedings of the international conference on information processing and management of uncertainty in knowledge-based systems 98, Paris
- Quinlan JR (1993) C4.5–programs for machine learning. Morgan Kaufmann, Palo Alto

- Shavlik JW, Dietterich TG (eds) (1990) Readings in machine learning. Morgan Kaufmann, Palo Alto
- Skowron A, Grzymala-Busse J (1994) From rough set theory to evidence theory. In: Yager R, Fedrizzi M, Kacprzyk J (eds) Advances in the Dempster-Shafer theory of evidence. Wiley, New York, pp 193–236
- Tsumoto S (1998a) Modelling medical diagnostic rules based on rough sets. In: Polkowski L, Skowron A (eds) Rough sets and current trends in computing. Lecture note in artificial intelligence, vol 1424. Springer, Berlin
- Tsumoto S (1998b) Automated extraction of medical expert system rules from clinical databases based on rough set theory. *J Inf Sci* 112:67–84
- Tsumoto S (1998c) Extraction of experts' decision rules from clinical databases using rough set model. *Intell Data Anal* 2(3):215–227
- Tsumoto S (2001) Medical diagnostic rules as upper approximation of rough sets. *Proc FUZZ-IEEE2001*
- Tsumoto S, Tanaka H (1996) Automated discovery of medical expert system rules from clinical databases based on rough sets. In: Proceedings of the 2nd international conference on knowledge discovery and data mining 96. AAAI Press, Palo Alto, pp 63–69
- Ziarko W (1993) Variable precision rough set model. *J Comput Syst Sci* 46:39–59

Rule Induction, Missing Attribute Values, and Discretization

Jerzy W. Grzymala-Busse

Department of Electrical Engineering and
Computer Science, University of Kansas,
Lawrence, KS, USA

Department of Expert Systems and Artificial
Intelligence, University of Information
Technology and Management, Rzeszow, Poland

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Discretization

LEM2 Algorithm

Inconsistent Data

Missing Attribute Values

MLEM2

Classification System

Validation

Future Directions

Bibliography

Glossary

Discretization

Discretization is a process of converting numerical attributes into symbolic ones by splitting the numerical attribute domain into intervals. Usually discretization is conducted before the main process of rule induction, but in some rule induction algorithms, e.g., in MLEM2 (Modified LEM2), rules are induced concurrently with discretization.

LEM2 Algorithm

LEM2 (Learning from Examples Module, version 2) is the basic rule induction algorithm of the

machine learning/data mining system LERS. LEM2, implemented for the first time in 1990, uses an idea of a local covering to induce minimal set of minimal rules describing all data concepts.

LERS Machine Learning/Data Mining System

LERS (Learning from Examples based on Rough Sets) is a rule induction system created at the University of Kansas. Its first implementation was done in Franz Lisp in 1988. This first version of LERS had only one algorithm called LEM1 (Learning from Examples Module, version 1) to induce all rules from input data.

Missing Attribute Values

Missing attribute values frequently affect real-life data. Some attribute values are *lost* (e.g., erased), some are of the type “*do not care*” *conditions* (such attribute values were irrelevant for classification of the case). In most existing machine learning/data mining systems, some method of handling missing attribute values is applied before the main process of rule induction. However, in MLEM2 rule induction and handling missing attribute values are conducted at the same time.

Rule Induction

Rule induction is understood here as an instance of supervised learning. Rule induction is one of the basic processes of acquiring knowledge (knowledge extraction) in the form of rule sets from raw data. This process is widely used in machine learning (data mining). A data set contains cases (examples) characterized by attribute values and classified as members of concepts by an expert. *Rules* are expressions of the following format.

if condition₁ **and** condition₂ **and**...**and**
condition_n **then** decision.

Definition of the Subject and Its Importance

Rule induction is a process of creating rule sets from raw data called *training data*. Such rules represent hidden and previously unknown knowledge contained in the training data. These rules may be used for successful classification of new cases that were not used for training. One of the possible applications of this methodology is rule-based expert systems. There are many documented examples of successful use of rule induction for medicine (e.g., for decision support for diagnosis), in finances, military, etc.

In particular, one of the rule induction systems called LERS has proven its applicability having been used for years by NASA Johnson Space Center (Automation and Robotics Division), as a tool to develop expert systems of the type used in medical decision-making on board the International Space Station. LERS was also used to enhance facility compliance under Sections 311, 312, and 313 of Title III, the Emergency Planning and Community Right to Know. System LERS was used in other areas as well, e.g., in the medical field to assess preterm labor risk for pregnant women. Currently used manual methods of assessing preterm birth have a positive predictive value of 17–38%. The data mining methods based on LERS reached positive predictive value of 59–92%. Other applications of LERS to medical area include diagnosis of melanoma, prediction of behavior under mental retardation and analysis of animal models for prediction of self-injurious behavior. LERS has been used also in nursing, global warming, environmental protection, natural language, and data transmission. LERS may process big datasets and frequently outperforms not only other rule induction systems but also human experts.

Introduction

Data mining uses methods of machine learning to acquire knowledge from raw data. Rule induction is one of the most successful techniques of machine learning. In many application areas,

such as medicine, it is essential not only to make an appropriate decision or diagnosis but also to be able to justify or explain the decision. Rules provide such explanations.

We will discuss rule induction using for that purpose one of the very successful algorithms called LEM2 (Learning from Examples Module, version 2), see, e.g., Grzymala-Busse (1992). This algorithm is used in the LERS (Learning from Examples based on Rough Sets) data mining system. LERS can process inconsistent data, i.e., data with conflicting cases, in which values for all attributes are the same yet the decision values are distinct. LERS deals with inconsistent data using typical rough set approach: it computes lower and upper approximations for all concepts. Then it passes such approximations to LEM2.

Data are frequently incomplete, i.e., some attribute values are missing. Rule sets may be induced from incomplete data using a modified LEM2 algorithm called MLEM2. Again, MLEM2 uses a rough set approach to missing attribute values.

Additionally, many real-life data sets contain numerical attributes, i.e., attributes whose values are integers or real numbers. Numerical data need to be converted into symbolic data by a process called discretization. For a numerical attribute, a few intervals are determined, usually these intervals are disjoint and together they cover the entire domain of the numerical attribute. The process of discretization may be performed before rule induction or, as it is done in MLEM2, during rule induction.

In the entire process of rule learning, handling missing attribute values, and discretization, the applied methodology is based on the same granules called *attribute-value blocks*. Similar granules were discussed in Lin (1989a, b, 1992).

Discretization

The input examples are presented in a table, called a decision table. An example of such a table is presented as Table 1. Table 1 represents the input data set in the format of LERS (Grzymala-Busse 1992). Table 1 contains four variables: *Age*, *Skill*, *Experience*, and *Productivity*. The first three variables are

Rule Induction, Missing Attribute Values, and Discretization,**Table 1** An example of a decision table

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25	low	low	low
2	56	high	high	high
3	36	low	high	low
4	42	high	low	low
5	59	high	low	high
6	25	high	low	low
7	42	high	high	high

Rule Induction, Missing Attribute Values, and Discretization,**Table 2** An example of an inconsistently discretized decision table

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25..42	low	low	low
2	42..57	high	high	high
3	25..42	low	high	low
4	42..57	high	low	low
5	42..57	high	low	high
6	25..42	high	low	low
7	42..57	high	high	high

called *attributes*, the last one is called *decision*. Table 1 presents seven cases (or examples) labeled by integers 1, 2, ..., 7.

Any subset of the set of all cases, defined by the same value of the decision is called a *concept*. Table 1 consists of two concepts, defined by (*Productivity*, *low*) and (*Productivity*, *high*). For example, the concept defined by (*Productivity*, *low*) is the set {1, 3, 4, 6}.

The attribute *Age* from Table 1 is numerical. The numerical attributes should be discretized before rule induction, otherwise rules induced from data with numerical attributes will be too specific. In general, during *discretization* an original range $[a, b]$ of a numerical attribute is partitioned into a set of n intervals.

$$\{[a_0, a_1), [a_1, a_2), \dots, [a_{n-1}, a_n]\},$$

where $a_0 = a$, $a_k < a_{k+1}$ for $k = 0, 1, \dots, n-1$ and $a_n = b$. The data mining system LERS uses for discretization a number of discretization algorithms (Chmielewski and Grzymala-Busse 1996),

including Equal Interval Width, Equal Frequency per Interval, Minimal Class Entropy, and two discretization methods based on Cluster Analysis (Grzymala-Busse 1997; Grzymala-Busse 2003b). As an example, let us use one of the simplest discretization methods, discretization based on the Equal Interval Width. Let us start with $n = 2$. Thus, the range $[25, 59]$ of *Age* is partitioned into two intervals of the equal width: $[25, 42)$ and $[42, 59]$. We will denote the first interval by 25..42 and the second interval by 42..59. As a result, the decision table that represents the discretized data set (Table 2) is inconsistent, since cases 4 and 5 are conflicting. Therefore, the number of discretization intervals was too small. Let us try $n = 3$. The new intervals are $[25, 36.3)$, $[36.3..47.7)$, and $[47.7..59]$, denoted, respectively, by 25..36.3, 36.3..47.7, and 47.7..59. The new discretized table is presented in Table 3. This table is consistent and may be considered to be an input data set for rule induction. For more details on discretization methods see Chmielewski and Grzymala-Busse (1996);

Grzymala-Busse (2002a, 2003b); Grzymala-Busse (2007).

LEM2 Algorithm

LEM2 explores the search space of attribute-value pairs. Its input data set is a lower or upper approximation of a concept, so its input data set is always consistent. In general, LEM2 computes a local covering and then converts it into a rule set. We will quote a few definitions to describe the LEM2 algorithm (Chan and Grzymala-Busse 1991; Grzymala-Busse 1992, 1997).

The LEM2 algorithm is based on an idea of an attribute-value pair block. For an attribute-value pair $(a, v) = t$, a *block* of t , denoted by $[t]$, is a set of all cases from U such that for attribute a have value v . Let B be a nonempty lower or upper approximation of a concept represented by a decision-value pair (d, w) . Set B *depends* on a set T of attribute-value pairs $t = (a, v)$ if and only if

$$\emptyset \neq [T] = \bigcap_{t \in T} [t] \subseteq B.$$

Set T is a *minimal complex* of B if and only if B depends on T and no proper subset T' of T exists such that B depends on T' . Let $\mathcal{T}$ be a nonempty collection of nonempty sets of attribute-value pairs. Then $\mathcal{T}$ is a *local covering* of B if and only if the following conditions are satisfied:

- Each member T of $\mathcal{T}$ is a minimal complex of B
- $\bigcup_{T \in \mathcal{T}} [T] = B$
- $\mathcal{T}$ is minimal, i.e., $\mathcal{T}$ has the smallest possible number of members

The procedure LEM2, based on rule induction from local coverings, is presented below.

Procedure LEM2

```

(input: a set  $B$ ,
output: a single local covering  $\mathcal{T}$  of set  $B$ );
begin
   $G := B$ ;
   $\mathcal{T} := \emptyset$ ;
  while  $G \neq \emptyset$ 
  begin
     $T := \emptyset$ ;
     $T(G) := \{t \mid [t] \cap G \neq \emptyset\}$ ;
    while  $T = \emptyset$  or  $[T] \not\subseteq B$ 
    begin
      select a pair  $t \in T(G)$  such that
         $[t] \cap G$  is maximum;
      if a tie occurs, select a pair
         $t \in T(G)$  with the smallest
        cardinality of  $[t]$ ;
      if another tie occurs,
        select first pair;
       $T := T \cup \{t\}$ ;
       $G := [t] \cap G$ ;
       $T(G) := \{t \mid [t] \cap G \neq \emptyset\}$ ;
       $T(G) := T(G) - T$ ;
    end {while}
    for each  $t \in T$  do
      if  $[T - \{t\}] \subseteq B$ 
      then  $T := T - \{t\}$ ;
     $\mathcal{T} := \mathcal{T} \cup \{T\}$ ;
     $G := B - \bigcup_{T \in \mathcal{T}} [T]$ ;
  end {while};
  for each  $T \in \mathcal{T}$  do
    if  $\bigcup_{S \in \mathcal{T} - \{T\}} [S] = B$  then  $\mathcal{T} := \mathcal{T} - \{T\}$ ;
  end {procedure}.

```

For a set X , $|X|$ denotes the cardinality of X .

For Table 3, the following rules are induced by LEM2:

Rule Induction, Missing Attribute Values, and Discretization,**Table 3** An example of a discretized decision table

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25..36.3	low	low	low
2	47.7..59	high	high	high
3	25..36.3	low	high	low
4	36.3..47.7	high	low	low
5	47.7..59	high	low	high
6	25..36.3	high	low	low
7	36.3..47.7	high	high	high

(Age, 25..36.3) $\rightarrow$ (Productivity, low),

(Age, 36.3..47.7) & (Experience, low) $\rightarrow$
(Productivity, low),

(Age, 47.7..59) $\rightarrow$ (Productivity, high),

(Age, 36.3..47.7) & (Experience, high) $\rightarrow$
(Productivity, high).

$$\{x \in U \mid [x]_P \subseteq X\}.$$

The second set is called an *upper approximation* of X in P , denoted by $\bar{P}X$ and defined as follows

$$\{x \in U \mid [x]_P \cap X \neq \emptyset\}.$$

Inconsistent Data

LEERS handles inconsistencies using rough set theory, introduced by Z. Pawlak in 1982 (Pawlak 1982, 1991). In rough set theory, lower and upper approximations are computed for all concepts involved in conflicts with other concepts.

Let U denote the set of all examples of the decision table and let P denote a nonempty subset of the set Q of all variables, i.e., attributes and decisions. Let P be a subset of A . An indiscernibility relation ρ on U is defined for all $x, y \in U$ by $x\rho y$ if and only if for both x and y the values for all variables from P are identical. Equivalence classes of ρ are called *elementary sets* of P . An equivalence class of ρ containing x is denoted $[x]_P$. Any finite union of elementary sets of P is called a *definable set* in P . Let X be a concept. In general, X is not a definable set in P . However, set X may be approximated by two definable sets in P , the first one is called a *lower approximation* of X in P , denoted by $\underline{P}X$ and defined as follows

The lower approximation of X in A is the greatest definable set in A , contained in X . The upper approximation of X in A is the least definable set in A containing X . A rough set of X is the family of all subsets of U having the same lower and the same upper approximations of X .

Table 4 is an example of inconsistent data. Cases 6 and 8 are conflicting. The lower approximation of the concept $\{1, 3, 4, 6\}$ is the set $\{1, 3, 4\}$ and the upper approximation of the same concept is the set $\{1, 3, 4, 6, 8\}$. The Table 4, discretized using the same Equal Interval Width method that was used before to Table 1, with three intervals, is presented as Table 5. The corresponding rule sets are: the certain rule set:

(Skill, low) $\rightarrow$ (Productivity, low)

(Age, 36.3..47.7) & (Experience, low) $\rightarrow$
(Productivity, low)

(Age, 47.7..59) $\rightarrow$ (Productivity, high),

(Age, 36.3..47.7) & (Experience, high) $\rightarrow$
(Productivity, high),

Rule Induction, Missing Attribute Values, and Discretization,

Table 4 An example of an inconsistent decision table

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25	low	low	low
2	56	high	high	high
3	36	low	high	low
4	42	high	low	low
5	59	high	low	high
6	25	high	low	low
7	42	high	high	high
8	25	high	low	high

Rule Induction, Missing Attribute Values, and Discretization,

Table 5 An example of a discretized and inconsistent decision table

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25..36.3	low	low	low
2	47.7..59	high	high	high
3	25..36.3	low	high	low
4	36.3..47.7	high	low	low
5	47.7..59	high	low	high
6	25..36.3	high	low	low
7	36.3..47.7	high	high	high
8	25..36.3	high	low	high

and the possible rule set:

(Age, 25..36.3) -> (Productivity, low),
 (Age, 36.3..47.7) & (Experience, low) ->
 (Productivity, low),
 (Skill, high) & (Age, 25..36.3) -> (Productivity,
 high),
 (Age, 47.7..59) -> (Productivity, high),
 (Age, 36.3..47.7) & (Experience, high) ->
 (Productivity, high).

Missing Attribute Values

For incomplete decision tables, there are three important and different possibilities to define lower and upper approximations, called singleton, subset, and concept approximations (Grzymala-Busse 2003a). Singleton lower and upper approximations were studied, e.g., in Kryszkiewicz (1995, 1999), Stefanowski and Tsoukias (1999,

2001). Note that similar definitions of lower and upper approximations, though not for incomplete decision tables, were studied in Lin (1992). Additionally, note that some other rough-set approaches to missing attribute values were presented in Grzymala-Busse (1992) as well.

For the rest of the paper we will assume that all decision values are specified, i.e., they are not missing. Also, we will assume that lost values will be denoted by “?” and “do not care” conditions by “*”. Additionally, we will assume that for each case at least one attribute value is specified.

For incomplete decision tables, the definition of a block of an attribute-value pair must be modified in the following way:

- If for an attribute a there exists a case x such that $\rho(x, a) = ?$, i.e., the corresponding value is lost, then the case x should not be included in any blocks $[(a, v)]$ for all values v of attribute a ,

- If for an attribute a there exists a case x such that the corresponding value is a “do not care” condition, i.e., $\rho(x, a) = *$, then the case x should be included in blocks $[(a, v)]$ for all specified values v of attribute a .

For Table 6 the blocks of attribute-value pairs are:

$[(\text{Age}, 25)] = \{1, 6\},$
 $[(\text{Age}, 36)] = \{3\},$
 $[(\text{Age}, 42)] = \{4, 7\},$
 $[(\text{Skill}, \text{low})] = \{1, 3, 4, 7\},$
 $[(\text{Skill}, \text{high})] = \{2, 4, 5, 6, 7\},$
 $[(\text{Experience}, \text{low})] = \{1, 2, 4, 5, 6\},$
 $[(\text{Experience}, \text{high})] = \{1, 2, 3, 6, 7\},$

For a case $x \in U$, the *characteristic set* $K_B(x)$ is defined as the intersection of the sets $K(x, a)$, for all $a \in B$, where the set $K(x, a)$ is defined in the following way:

- If $\rho(x, a)$ is specified, then $K(x, a)$ is the block $[(a, \rho(x, a))]$ of attribute a and its value $\rho(x, a)$,
- If $\rho(x, a) = ?$ or $\rho(x, a) = *$ then the set $K(x, a) = U$.

For Table 6 and $B = A$,

$K_A(1) = \{1, 6\} \cap \{1, 3, 4, 7\} = \{1\},$
 $K_A(2) = \{2, 4, 5, 6, 7\},$

$K_A(3) = \{3\} \cap \{1, 3, 4, 7\} \cap \{1, 2, 3, 6, 7\} = \{3\},$
 $K_A(4) = \{4, 7\} \cap \{1, 2, 4, 5, 6\} = \{4\},$
 $K_A(5) = \{2, 4, 5, 6, 7\} \cap \{1, 2, 4, 5, 6\} = \{2, 4, 5, 6\},$
 $K_A(6) = \{1, 6\} \cap \{2, 4, 5, 6, 7\} = \{6\},$
 $K_A(7) = \{4, 7\} \cap \{1, 2, 3, 6, 7\} = \{7\}.$

Characteristic set $K_B(x)$ may be interpreted as the set of cases that are indistinguishable from x using all attributes from B and using a given interpretation of missing attribute values.

Lower and Upper Approximations

For incomplete decision tables, lower and upper approximations may be defined in a few different ways. Here we suggest three different definitions of lower and upper approximations for incomplete decision tables, following Grzymala-Busse (2003a). Let B be a subset of the set A of all attributes. Let X be any subset of the set U of all cases. The set X is called a *concept* and is usually defined as the set of all cases defined by a specific value of the decision.

Our first definition uses a similar idea as in the previous articles on incomplete decision tables (Kryszkiewicz 1995, 1999; Stefanowski and Tsoukias 1999, 2001), i.e., lower and upper approximations are sets of singletons from the universe U satisfying some properties. We will call these approximations *singleton*. A singleton B -lower approximation of X is defined as follows:

Rule Induction, Missing Attribute Values, and Discretization,

Table 6 An example of a decision table with missing attribute values

Case	Attributes			Decision
	Age	Skill	Experience	Productivity
1	25	low	*	low
2	?	high	*	high
3	36	low	high	low
4	42	*	low	low
5	?	high	low	high
6	25	high	*	low
7	42	*	high	high

$$\underline{B}X = \{x \in U \mid K_B\{x\} \subseteq X\}.$$

A singleton B -upper approximation of X is

$$\overline{B}X = \{x \in U \mid K_B(x) \cap X \neq \emptyset\}.$$

In our example of the decision table presented in Table 6, let us say that $B = A$. Then the singleton A -lower and A -upper approximations of the two concepts: $\{1, 3, 4, 6\}$ and $\{2, 5, 7\}$ are:

$$\begin{aligned}\underline{A}\{1, 3, 4, 6\} &= \{1, 3, 4, 6\}, \\ \underline{A}\{2, 5, 7\} &= \{7\}, \\ \overline{A}\{1, 3, 4, 6\} &= \{1, 2, 3, 4, 5, 6\}, \\ \overline{A}\{2, 5, 7\} &= \{2, 5, 7\}.\end{aligned}$$

The next possibility is to define lower and upper approximations for incomplete decision tables using characteristic sets instead of singletons. There are two ways to do this. Using the first way, a *subset* B -lower approximation of X is defined as follows:

$$\underline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \subseteq X\}.$$

A *subset* B -upper approximation of X is

$$\overline{B}X = \cup\{K_B(x) \mid x \in U, K_B(x) \cap X \neq \emptyset\}.$$

For the same decision table, presented in Table 6, the subset A -lower and A -upper approximations are

$$\begin{aligned}\underline{A}\{1, 3, 4, 6\} &= \{1, 3, 4, 6\}, \\ \underline{A}\{2, 5, 7\} &= \{7\}, \\ \overline{A}\{1, 3, 4, 6\} &= \{1, 2, 3, 4, 5, 6, 7\}, \\ \overline{A}\{2, 5, 7\} &= \{2, 4, 5, 6, 7\}.\end{aligned}$$

The second way is to modify the subset definition of lower and upper approximation by replacing the universe U from the subset definition by a concept X . A *concept* B -lower approximation of the concept X is defined as follows:

$$\underline{B}X = \cup\{K_B(x) \mid x \in X, K_B(x) \subseteq X\}.$$

A *concept* B -upper approximation of the concept X is defined as follows:

$$\begin{aligned}\overline{B}X &= \cup\{K_B(x) \mid x \in X, K_B(x) \cap X \neq \emptyset\} \\ &= \cup\{K_B(x) \mid x \in X\}.\end{aligned}$$

For the decision table presented in Table 6, the concept A -upper approximations are

$$\begin{aligned}\overline{A}\{1, 3, 4, 6\} &= \{1, 3, 4, 6\}, \\ \overline{A}\{2, 5, 7\} &= \{2, 4, 5, 6, 7\}.\end{aligned}$$

Note that for complete decision tables, all three definitions of lower approximations, singleton, subset, and concept, coalesce to the same definition. Also, for complete decision tables, all three definitions of upper approximations coalesce to the same definition. This is not true for incomplete decision tables, as our example shows.

Probabilistic Approximations

Three kinds of probabilistic approximations, called singleton, subset, and concept, were introduced in Clark and Grzymala-Busse (2011). As it was shown in Clark and Grzymala-Busse (2011) and Clark et al. (2014), probabilistic approximations are useful for mining incomplete data sets.

A B -singleton probabilistic approximation of X with the threshold α , $0 < \alpha \leq 1$, denoted by $\text{appr}_{\alpha, B}^{\text{singleton}}(X)$, is defined by

$$\{x \mid x \in U, \text{Pr}\{X \mid K_B(x)\} \geq \alpha\},$$

where $\text{Pr}(X \mid K_B(x)) = \frac{|X \cap K_B(x)|}{|K_B(x)|}$ is the conditional probability of X given $K_B(x)$ and $|Y|$ denotes the cardinality of set Y . A B -subset probabilistic approximation of the set X with the threshold α , $0 < \alpha \leq 1$, denoted by $\text{appr}_{\alpha, B}^{\text{subset}}(X)$, is defined by

$$\cup\{K_B(x) \mid x \in U, \text{Pr}\{X \mid K_B(x)\} \geq \alpha\}.$$

A B -concept probabilistic approximation of the set X with the threshold α , $0 < \alpha \leq 1$, denoted by $\text{appr}_{\alpha, B}^{\text{concept}}(X)$, is defined by

$$\cup\{K_B(x) \mid x \in X, \text{Pr}(X \mid K_B(x)) \geq \alpha\}.$$

Note that lower and upper approximations are special cases for probabilistic approximations. A probabilistic approximation for $\alpha = 1$ is the lower approximation of the same type; similarly, a probabilistic approximation with the smallest positive α is the upper approximation of the same type.

For Table 6, all distinct probabilistic approximations are

$$appr_{0.4,A}^{singleton}(\{1, 3, 4, 6\}) = \{1, 2, 3, 4, 5, 6\},$$

$$appr_{0.5,A}^{singleton}(\{1, 3, 4, 6\}) = \{1, 3, 4, 5, 6\},$$

$$appr_{1,A}^{singleton}(\{1, 3, 4, 6\}) = \{1, 3, 4, 6\},$$

$$appr_{0.5,A}^{singleton}(\{2, 5, 7\}) = \{2, 5, 7\},$$

$$appr_{0.6,A}^{singleton}(\{2, 5, 7\}) = \{2, 7\},$$

$$appr_{1,A}^{singleton}(\{2, 5, 7\}) = \{7\},$$

$$appr_{0.4,A}^{subset}(\{1, 3, 4, 6\}) = \{1, 2, 3, 4, 5, 6, 7\},$$

$$appr_{0.5,A}^{subset}(\{1, 3, 4, 6\}) = \{1, 2, 3, 4, 5, 6\},$$

$$appr_{1,A}^{subset}(\{1, 3, 4, 6\}) = \{1, 3, 4, 6\},$$

$$appr_{0.6,A}^{subset}(\{2, 5, 7\}) = \{2, 4, 5, 6, 7\},$$

$$appr_{1,A}^{subset}(\{2, 5, 7\}) = \{7\},$$

$$appr_{1,A}^{concept}(\{1, 3, 4, 6\}) = \{1, 3, 4, 6\},$$

$$appr_{0.6,A}^{concept}(\{2, 5, 7\}) = \{2, 4, 5, 6, 7\},$$

$$appr_{1,A}^{concept}(\{2, 5, 7\}) = \{7\}.$$

If for some β , $0 < \beta \leq 1$, a probabilistic approximation $appr_{\beta}^{CS}(X)$ is not listed above, it is equal to the probabilistic approximation $appr_{\alpha}^{CS}(X)$ with the closest α to β , $\alpha \geq \beta$. For example, $appr_{0.45,A}^{singleton}(\{1, 3, 4, 6\}) = appr_{0.5,A}^{singleton}(\{1, 3, 4, 6\})$.

MLEM2

MLEM2, a modified version of LEM2 (Grzymala-Busse 2002b), processes numerical attributes differently than symbolic attributes. For numerical attributes MLEM2 sorts all values of a numerical attribute. Then it computes cut points as averages for any two consecutive values of the sorted list. For each cutpoint q MLEM2 creates two blocks, the first block contains all cases for which values of the numerical attribute are smaller than q , the second block contains remaining cases, i.e., all cases for which values of the numerical attribute are larger than q . The search space of MLEM2 is the set of all blocks computed this way, together with blocks defined by symbolic attributes. Starting from that point, rule induction in MLEM2 is conducted the same way as in LEM2. At the very end, MLEM2 simplifies rules by merging appropriate intervals for numerical attributes.

The MLEM algorithm induced the following rules from Table 6: certain rule.

set:

(Age, 25..39) -> (Productivity, low),
 (Skill, low) & (Experience, low) -> (Productivity, low),
 (Age, 39..42) & (Experience, high) -> (Productivity, high),

and possible rule set:

(Age, 25..39) -> (Productivity, low),
 (Skill, low) & (Experience, low) -> (Productivity, low),
 (Skill, high) -> (Productivity, high).

Classification System

Rule sets, induced from data sets, are used mostly to classify new, unseen cases. Such rule sets may be used in rule-based expert systems.

There are a few existing classification systems, e.g., associated with rule induction systems LERS or AQ. A classification system used in LERS is a

modification of the well-known bucket brigade algorithm (Grzymala-Busse 1997; Stefanowski 2001). In the rule induction system AQ, the classification system is based on a rule estimate of probability. Some classification systems use a decision list, in which rules are ordered, the first rule that matches the case classifies it. In this section, we will concentrate on a classification system used in LERS.

The decision to which concept a case belongs to is made on the basis of three factors: *strength*, *specificity*, and *support*. These factors are defined as follows: *strength* is the total number of cases correctly classified by the rule during training. *Specificity* is the total number of attribute-value pairs on the left-hand side of the rule. The matching rules with a larger number of attribute-value pairs are considered more specific. The third factor, *support*, is defined as the sum of products of strength and specificity for all matching rules indicating the same concept. The concept C for which the support, i.e., the following expression

$$\sum_{\text{matching rules } r \text{ describing } C} \text{Strength}(r) * \text{Specificity}(r)$$

is the largest is the winner and the case is classified as being a member of C .

In the classification system of LERS, if complete matching is impossible, all partially matching rules are identified. These are rules with at least one attribute-value pair matching the corresponding attribute-value pair of a case. For any partially matching rule r , the additional factor, called *Matching_factor* (r), is computed. *Matching_factor* (r) is defined as the ratio of the number of matched attribute-value pairs of r with a case to the total number of attribute-value pairs of r . In partial matching, the concept C for which the following expression is the largest

$$\sum_{\substack{\text{partially matching} \\ \text{rules } r \text{ describing } C}} \text{Matching_factor}(r) * \text{Strength}(r) * \text{Specificity}(r)$$

is the winner and the case is classified as being a member of C .

Validation

The most important performance criterion of rule induction methods is the error rate. If the number of cases is less than 100, the *leaving-one-out* method is used to estimate the error rate of the rule set. In leaving-one-out, the number of learn-and-test experiments is equal to the number of cases in the data set. During the i -th experiment, the i -th case is removed from the data set, a rule set is induced by the rule induction system from the remaining cases, and the classification of the omitted case by rules produced is recorded. The error rate is computed as

$$\frac{\text{total number of misclassifications}}{\text{total number of cases}}.$$

On the other hand, if the number of cases in the data set is greater than or equal to 100, the *tenfold cross-validation* should be used. This technique is similar to leaving-one-out in that it follows the learn-and-test paradigm. In this case, however, all cases are randomly re-ordered, and then a set of all cases is divided into ten mutually disjoint subsets of approximately equal size. For each subset, all remaining cases are used for training, i.e., for rule induction, while the subset is used for testing. This method is used primarily to save time at the negligible expense of accuracy.

Tenfold cross-validation is commonly accepted as a standard way of validating rule sets. However, using this method twice, with different preliminary random reordering of all cases, yields – in general – two different estimates for the error rate (Grzymala-Busse 1997). For large data sets (at least 1000 cases), a single application of the train-and-test paradigm may be used. This technique is also known as *hold-out*. Two thirds of cases should be used for training, one third for testing.

In yet another way of validation, *resubstitution*, it is assumed that the training data set is identical with the testing data set. In general, an estimate for the error rate is here too optimistic. However, this technique is used in many applications.

Future Directions

Rule induction, a part of supervised learning, is a subject to extensive research. Many possible extensions for rule induction include *boosting* or *ensemble learning*, where ensembles of classifiers vote on the final outcome during classification of a new case. *Semi-supervised learning*, where for learning is used not only the case set, preclassified by an expert, but also a set of not classified cases, is another example.

Bibliography

- Chan CC, Grzymala-Busse JW (1991) On the attribute redundancy and the learning programs ID3, PRISM, and LEM2. Department of Computer Science, University of Kansas, TR-91-14, 20 p
- Chmielewski MR, Grzymala-Busse JW (1996) Global discretization of continuous attributes as preprocessing for machine learning. *Int J Approx Reason* 15:319–331
- Clark PG, Grzymala-Busse JW (2011) Experiments on probabilistic approximations. In: Proceedings of the IEEE international conference on granular computing, Kaohsiung, pp 144–149
- Clark PG, Grzymala-Busse JW, Rzas W (2014) Mining incomplete data with singleton, subset and concept approximations. *Inf Sci* 280:368–384
- Grzymala-Busse JW (1988) Knowledge acquisition under uncertainty – a rough set approach. *J Intell Robot Syst* 1:3–16
- Grzymala-Busse JW (1992) LERS – A system for learning from examples based on rough sets. In: Slowinski R (ed) *Intelligent decision support. Handbook of applications and advances of the rough set theory*. Kluwer, Dordrecht/Boston/London, pp 3–18
- Grzymala-Busse JW (1997) A new version of the rule induction system LERS. *Fund Inform* 31:27–39
- Grzymala-Busse JW (2002a) Discretization of numerical attributes. In: Klösgen W, Zytkow J (eds) *Handbook of data mining and knowledge discovery*. Oxford University Press, New York, pp 218–225
- Grzymala-Busse JW (2002b) MLEM2: a new algorithm for rule induction from imperfect data. In: Proceedings of the 9th international conference on information processing and management of uncertainty in knowledge-based systems, Annecy, pp 243–250
- Grzymala-Busse JW (2003a) Rough set strategies to data with missing attribute values. In: Workshop notes, foundations and new directions of data mining, in conjunction with the 3rd IEEE international conference on data mining, Melbourne, pp 56–63
- Grzymala-Busse JW (2003b) A comparison of three strategies to rule induction from data with numerical attributes. In: Proceedings of the international workshop on rough sets in knowledge discovery, in conjunction with the European joint conferences on theory and practice of Software, Warsaw, pp 132–140
- Grzymala-Busse JW (2007) Mining numerical data – a rough set approach. In: Proceedings of the international conference of rough sets and emerging intelligent systems paradigms, Warsaw, Poland. Lecture notes in AI, vol 4585. Springer, Berlin/Heidelberg/New York, pp 12–21
- Kryszkiewicz M (1995) Rough set approach to incomplete information systems. In: Proceedings of the second annual joint conference on information sciences, pp 194–197
- Kryszkiewicz M (1999) Rules in incomplete information systems. *Inf Sci* 113:271–292
- Lin TY (1989a) Neighborhood systems and approximation in database and knowledge base systems. In: Proceedings the fourth international symposium on methodologies of intelligent systems, Charlotte, pp 75–86
- Lin TY (1989b) Chinese Wall security policy – an aggressive model. In: Proceedings of the fifth aerospace computer security application conference, Tucson, pp 286–293
- Lin TY (1992) Topological and fuzzy rough sets. In: Slowinski R (ed) *Intelligent decision support. Handbook of applications and advances of the rough set theory*. Kluwer, Dordrecht/Boston/London, pp 287–304
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11:341–356
- Pawlak Z (1991) Rough sets. Theoretical aspects of reasoning about data. Kluwer, Dordrecht/Boston/London
- Stefanowski J (2001) Algorithms of decision rule induction in data mining. Poznan University of Technology Press, Poznan
- Stefanowski J, Tsoukias A (1999) On the extension of rough sets under incomplete information. In: Proceedings 7th international workshop on new directions in rough sets, data mining, and granular-soft computing, pp 73–81
- Stefanowski J, Tsoukias A (2001) Incomplete information tables and rough classification. *Comput Intell* 17:545–566

Social Networks and Granular Computing

Churn-Jung Liao

Institute of Information Science, Academia
Sinica, Taipei, Taiwan

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Mathematical Preliminaries

Rough Set Theory: Functional Granulation

Positional Analysis: Relational Granulation

Modal Logic: The Bridge

From Relational Granulation to Functional
Granulation

Related Work

Conclusion and Future Directions

Bibliography

Glossary

Exact equivalence Two actors are exactly equivalent if they are related to the same number of equivalent actors.

Functional granulation The granulation process is functional if it is based on the attributes of the objects. It is called functional granulation because each attribute is a function from the set of objects to the set of values.

Granulation The process of drawing a set of objects (or points) together by indiscernibility, similarity, proximity, or functionality.

Hybrid logic Hybrid logic is a branch of modal logic that allows direct reference to the elements in a relational structure. Traditionally, only the properties of the elements could be represented by modal logic formulas.

Modal logic Modal logic was originally developed as a type of philosophical logic for

reasoning about necessity and possibility. However, it has been extended to broadly cover a family of logics for reasoning about modalities including tense, obligation, belief, and knowledge. Semantically, it is also a powerful mathematical discipline that deals with (restricted) description languages for discussing various kinds of relational structures, where a relational structure comprises a set of elements and a collection of relations on that set.

Positional equivalence Two actors are in equivalent positions if their “pattern” of relationships with other actors is the same.

Regular equivalence Two actors are regularly equivalent if they are equally related to equivalent actors.

Relational granulation The granulation process is relational if it is based on the relationships between objects.

Rough set A rough set is defined by the lower and upper approximations of a concept. The lower approximation contains all elements that necessarily belong to the concept, while the upper approximation contains those that possibly belong to the concept. In rough set theory, a concept is considered a classical set.

Social network A social network is comprised of a set of actors, called the domain, and a family of relations on the domain. It is usually represented as a graph, where each node represents an actor and an edge between two nodes represents a relational tie between these two actors. An edge can be labeled with the relation it represents.

Structural equivalence Two actors are structurally equivalent if they are related to the same actors.

Definition of the Subject and Its Importance

Granular computing (GrC) is a problem-solving concept deeply rooted in human thinking. Hence, it has played a major role in solving many

important problems throughout the history of mathematics. GrC is concerned with the processing of information granules, which are groups of objects drawn together by indiscernibility, similarity, proximity, or functionality (Zadeh 1997). The process of forming information granules is called granulation. If the process is based totally on the attributes of the objects, it is called *functional granulation*, since attributes are mathematical functions from the set of objects to the set of values; if, in addition, the granulation process is also based on the relationship between objects, it is called *relational granulation* (Fan et al. 2006; Liao and Lin 2005) (Fig. 1).

Interestingly, social scientists have applied the techniques of relational granulation (albeit by different names) to positional analysis in social networks (Borgatti and Everett 1989; Boyd and Everett 1999; Everett and Borgatti 1994; Lerner 2005; White and Reitz 1983). Social network analysis (SNA) is a methodology used extensively in social and behavioral sciences, as well as in political science, economics, organization theory, and industrial engineering (Hanneman and Riddle 2005; Scott 2000; Wasserman and Faust 1994). Positional analysis of a social network tries to find similarities between actors in the network. While many

traditional clustering methods are based on the attributes of the individual actors, SNA is more concerned with the structural similarity between the actors. In SNA, a category, called a *social role* or *social position*, is defined in terms of the similarities of the patterns of relations among the actors, rather than the attributes of the actors (Fig. 2).

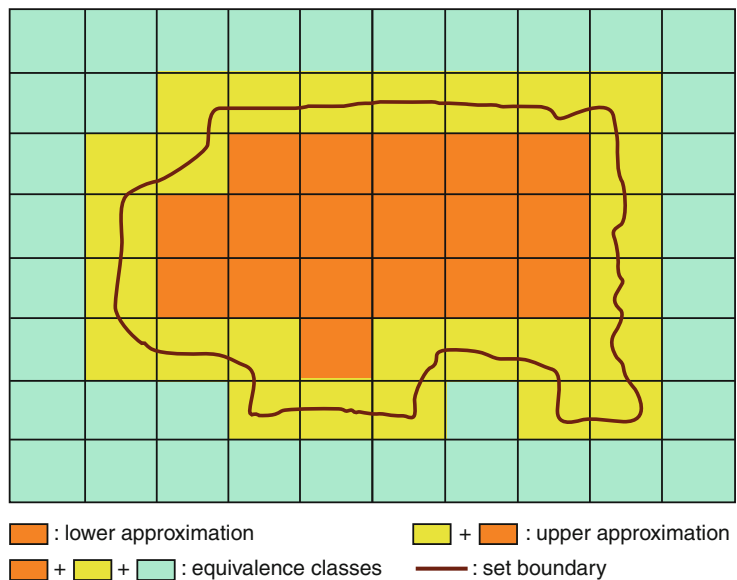
Analyzing a social position from a GrC perspective has two major advantages. On the one hand, it extends the application scope of the GrC methodology. On the other hand, we can logically transform relational granulation of social networks into functional granulation; as a result, the well-developed techniques of data table analysis can be applied to social position analysis. Thus, this is an important cross-fertilization of GrC and SNA.

Introduction

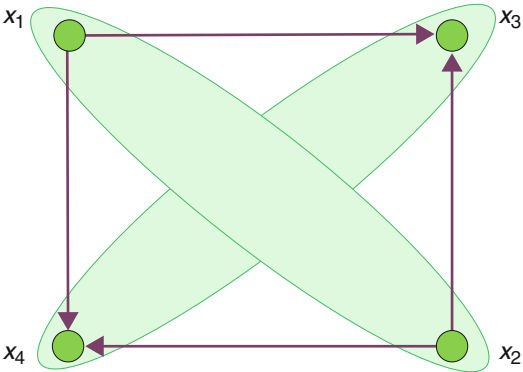
GrC encompasses a wide range of techniques in computational intelligence, such as interval computing, classification, cluster analysis, and fuzzy and rough set theories. Among them, rough set theory (Pawlak 1982, 1991) is one of the most well-known methods with successful applications in data analysis and uncertainty management. The

Social Networks and Granular Computing,

Fig. 1 A rough set: the set delimited by the curve (the set boundary) cannot be represented exactly as a union of the basic building blocks. The blocks totally included in the set form the lower approximation of the set, while the blocks that intersect with the set (including the lower approximation) form the upper approximation of the set



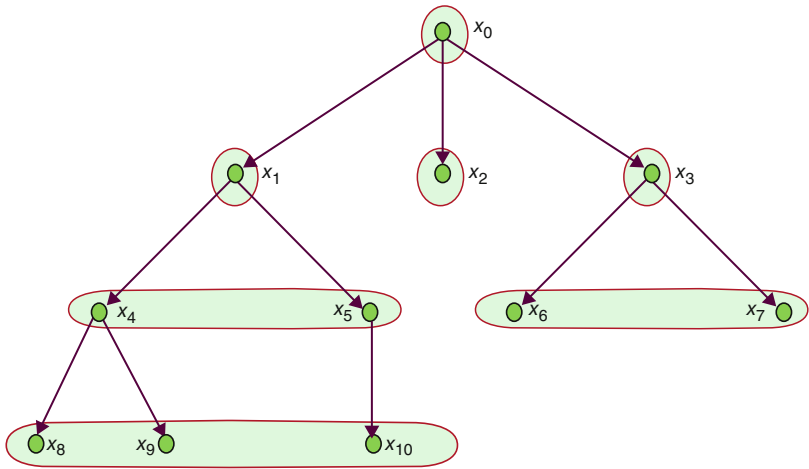
theory was first proposed by Z. Pawlak as a mathematical tool to deal with vague concepts (Pawlak 1982). The basic assumption of the theory is that the effective use of knowledge is based on the



Social Networks and Granular Computing, Fig. 2 Structural equivalence: the social network contains four actors denoted by the nodes and a single binary relation denoted by the arcs. Both the out-neighborhoods of x_1 and x_2 are $\{x_3, x_4\}$, and their common in-neighborhood is the empty set. On the other hand, both the in-neighborhoods of x_3 and x_4 are $\{x_1, x_2\}$, and their common out-neighborhood is the empty set. Thus, $x_1 \cong_s x_2$ and $x_3 \cong_s x_4$. The green shadowed areas indicate the equivalence classes

capability to classify objects. Thus, knowledge consists of a family of classifications of a domain of interest. A classification of a domain is represented as a partition of (or an equivalence relation on) the domain. Each equivalence class of a classification is the elementary knowledge about the classification. For example, if the domain is classified according to colors, then the elementary knowledge may include “red,” “green,” “blue,” etc. In this sense, a classification is derivable from a set of features or attributes of objects. An important task of data analysis is to define a concept via a set of features. If the concept can be precisely defined as the objects with some specific attribute values, then it is an exact concept with respect to the given attributes. However, in practice, most concepts are vague and cannot be defined precisely by a given set of attributes. In such cases, rough set theory provides an effective way to approximate the vague concepts with some exact concepts in the domain. Because of its applications in data mining, rough set theory has flourished since 1990 (Fig. 3).

At about the time rough set theory was proposed, the pioneer of fuzzy set theory, L.A. Zadeh,



Social Networks and Granular Computing, Fig. 3 Regular equivalence in a tree-structured social network: the root x_0 is discernible from other nodes because its in-neighborhood is empty, whereas all the other nodes have nonempty in-neighborhoods. Nodes x_1 , x_2 , and x_3 are discernible from other nodes because they are the only nodes connected to x_0 . Moreover, they are mutually discernible, since x_2 has no children, x_1 has children

and grandchildren, and x_3 has children, but not grandchildren. Note that x_4 and x_5 are indiscernible, since they have the same in-neighborhood and equivalent out-neighborhoods; x_6 and x_7 are indiscernible, since they have the same in-neighborhood and both have empty out-neighborhoods; and x_8 , x_9 , and x_{10} are indiscernible, since they have equivalent in-neighborhoods and all have empty out-neighborhoods

emphasized the importance of information granularity in fuzzy reasoning (Zadeh 1979). Subsequently, T.Y. Lin proposed the term “GrC” to unify different techniques developed for processing information granules. From the GrC perspective, both classification in rough set theory and fuzzification in fuzzy set theory are types of granulation.

In rough set theory, objects are partitioned into equivalence classes based on their attribute values. While such values essentially represent functional information associated with the objects, more complicated information can be utilized in the classification of objects. In particular, relational information between objects can play an important role in the granulation process. The information is defined by general binary relations, which are extensions of the functional attributes of the objects. Geometrically, such granulation is derived from the neighborhood system of topological spaces (Sierpinski and Krieger 1956), where each point/object is assigned at most one neighborhood/granule. This kind of granulation is called *relational granulation*, while granulation based on attribute values only is called *functional granulation* (Fan et al. 2006; Liao and Lin 2005).

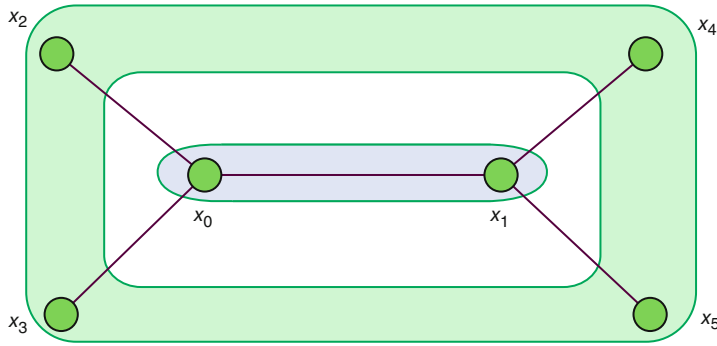
Relational granulation is especially useful for the analysis of network data because a network is a collection of relations between nodes. A concrete instance of the application of relational granulation is positional analysis of social networks. Positional analysis attempts to find actors occupying the same position in a social network based on the pattern of their relationships to other actors. Depending on the different patterns of relationships, different notions of positional equivalence have been proposed (Lerner 2005). These notions can be easily regarded as instances of relational granulation. In this article, we consider three of the most important notions of positional equivalence – *structural equivalence*, *regular equivalence*, and *exact equivalence*.

Recently, it was shown that social positions based on regular equivalence can be syntactically expressed as well-formed formulas (wffs) in a kind of modal logic (Marx and Masuch 2003). Thus, actors occupying the same social position based on regular equivalence will satisfy the same

set of modal formulas. Traditionally, modal logic has been considered the logic for reasoning about modalities, such as necessity, possibility, time, action, belief, knowledge, and obligation. However, semantically, it is essentially a language for describing relational structures (Blackburn et al. 2001). A relational structure is simply a set combined with a collection of relations on that set. Thus, social networks are mathematically equivalent to relational structures. The logical characterization of social positions implies that relational granulation for positional analysis can be transformed into a functional granulation process. Indeed, since each actor in a social network may satisfy or falsify a modal formula, a modal formula can be seen as a binary attribute associated with each actor. Therefore, based on the results in Marx and Masuch (2003), two actors are regularly equivalent if they have the same values with respect to all attributes defined by the language of the particular modal logic (Fig. 4).

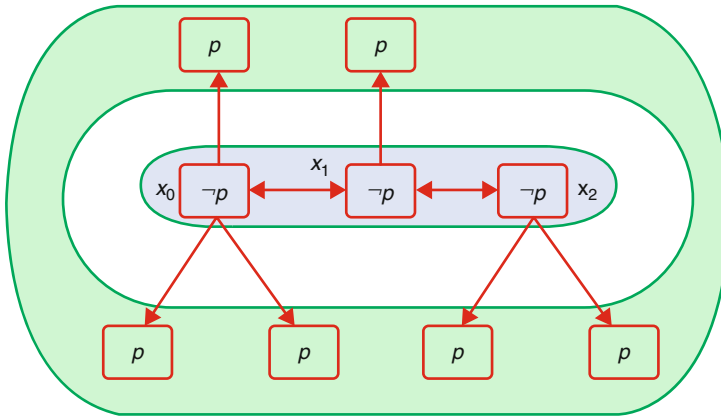
By extending previous results, we can find logical characterizations of different positional equivalences and transform them into a functional granulation process. Consequently, the techniques of rough set-based data analysis can also be applied to positional analysis. However, since the set of wffs of a modal logic is usually infinite, the functional granulations corresponding to positional equivalences need to consider infinite sets. To circumvent this problem, we also propose a new definition of positional equivalence, called *observational equivalence*. Let Σ be a set of wffs in a logical language that denotes the set of observable properties of the actors in a network. Two actors a and b are observationally equivalent with respect to Σ , if for each wff $\varphi \in \Sigma$, a satisfies φ iff b satisfies φ .

The remainder of this article is organized as follows. In section “[Mathematical Preliminaries](#),” we introduce the background knowledge and mathematical notations of sets and relations. In sections “[Rough Set Theory: Functional Granulation](#),” “[Positional Analysis: Relational Granulation](#),” and “[Modal Logic: The Bridge](#),” we review rough set theory, social positional analysis, and modal logic, respectively. In section “[From Relational Granulation to Functional](#)



Social Networks and Granular Computing, Fig. 4 Exact equivalence: the social network is an undirected graph. In other words, the binary relation is symmetric. The exact equivalence partitions the nodes into two

classes, $C_0 = \{x_0, x_1\}$ and $C_1 = \{x_2, x_3, x_4, x_5\}$. To verify that it is indeed an exact equivalence, we can check that both x_0 and x_1 connect to one C_0 node and two C_1 nodes, and all of x_2, x_3, x_4 , and x_5 connect to one C_0 node



Social Networks and Granular Computing, Fig. 5 A social network represented as a Kripke model: we assume the underlying modal language contains one propositional symbol p and one relational symbol r . Since $\{x_0, x_1, x_2\}$ is an equivalence class in the shown regular equivalence, the

three nodes all satisfy the same set of PMML wffs. For example, they all satisfy $\neg p \wedge \langle r \rangle p$. However, they are not exactly equivalent, so they are distinguishable by GML wffs. For example, x_0 satisfies both $\langle r \rangle_1 p$ and $\langle r \rangle_2 p$, x_1 does not satisfy $\langle r \rangle_1 p$ or $\langle r \rangle_2 p$, and x_2 satisfies $\langle r \rangle_1 p$ but not $\langle r \rangle_2 p$

[Granulation](#),” we present the logical characterizations of three main positional equivalences and the definition of observational equivalence. Finally, in section “[Conclusion and Future Directions](#),” we present our conclusions and indicate some future research directions (Fig. 5).

Mathematical Preliminaries

In this section, we introduce the basic knowledge about sets and relations used in this article.

Sets

A set is a collection of objects, called the elements of the set. We use capitals A, B, U, X , etc. to denote sets. The elements of a set are denoted by lower-case letters, and we write $x \in U$ if x is an element of the set U . We usually use two notations to denote a set. The first one lists its elements. For example, we can write $\{x_1, x_2, \dots, x_n\}$ for a finite set, where n is a positive integer, and $\{x_1, x_2, \dots\}$ for an infinite set. The second notation specifies the defining property of the set. For example, we can write $\{x \mid x = 2n, n \in \mathbb{Z}\}$ for the set of even numbers. The number of elements in a set is called

its cardinality. The cardinality of U is denoted by $|U|$. A set U is called an empty set if $|U| = 0$ and a singleton if $|U| = 1$. We denote the empty set by $\emptyset$.

The basic operations on sets are defined as follows:

1. Intersection of X and Y : $X \cap Y = \{u \mid u \in X \text{ and } u \in Y\}$.
2. Union of X and Y : $X \cup Y = \{u \mid u \in X \text{ and } u \in Y\}$.
3. Cartesian product of X and Y : $X \times Y = \{(x, y) \mid x \in X \text{ or } y \in Y\}$.
4. Difference of X by Y : $X - Y = \{u \mid u \in X \text{ and } u \notin Y\}$.

Since intersection, union, and the Cartesian product are associative (i.e., $(X \cap Y) \cap Z = X \cap (Y \cap Z)$, etc.), we can eliminate the parentheses and apply the operations to more than two sets. For example, the Cartesian product of the sets $U_1, U_2, \dots, U_k$ is denoted by

$$U_1 \times U_2 \times \dots \times U_k \\ = \{(x_1, x_2, \dots, x_k) \mid x_i \in U_i, \forall 1 \leq i \leq k\},$$

where each element is called a k -tuple. A 2-tuple is also called a pair. If $U_1 = U_2 = \dots = U_k = U$, the Cartesian product is written as U^k . A set X is said to be a subset of another set Y , denoted by $X \subseteq Y$, if for all $x \in X$, it is also the case that $x \in Y$. A function f from the set X to the set Y , denoted by $f: X \rightarrow Y$, is a mapping that associates each element in X with an element in Y . The set X is called the domain of the function, and Y is called the range of f . For each $x \in X$, we write $f(x)$ for the element in Y that is associated with x by f and call it the image of x under f . Also, the image of X under f is defined as $f(X) = \{f(x) \mid x \in X\}$.

Multisets

For sets, repeat occurrences of the same elements are only counted once. Thus, the set $\{x, x\}$ is the same as the set $\{x\}$. Sometimes, the number of occurrences of an element is important. In such cases, we consider multisets (or bags). A multiset is formally defined as a pair $M = (U, m)$, where U is some set and $m: U \rightarrow N$ is a function from

U to the set $N = \{0, 1, 2, 3, \dots\}$ of natural numbers. U is called the universe, and for each element $x \in U$, $m(x)$ is the multiplicity (i.e., the number of occurrences) of x in M . A set can be considered as a special multiset in which $m(x) = 0$ or 1 for each x in the universe. By generalizing the notation of set membership, we can write $x \in M$ if $m(x) > 0$. The cardinality of M is defined as $|M| = \sum_{x \in U} m(x)$. Let $M_1 = (U, m_1)$ and $M_2 = (U, m_2)$ be two multisets in the same universe; then, $M_1 \subseteq M_2$ iff $m_1(x) \leq m_2(x)$ for all $x \in U$. The intersection, union, and difference between $M_1 = (U, m_1)$ and $M_2 = (U, m_2)$ are defined as follows:

1. $M_1 \cap M_2 = (U, m)$, where $m(x) = \min(m_1(x), m_2(x))$ for all $x \in U$.
2. $M_1 \cup M_2 = (U, m)$, where $m(x) = \max(m_1(x), m_2(x))$ for all $x \in U$.
3. $M_1 - M_2 = (U, m)$, where $m(x) = (m_1(x) - m_2(x)) \cdot \chi(m_1(x) - m_2(x))$ for all $x \in U$, and

$$\chi(k) = \begin{cases} 1, & \text{if } k > 0, \\ 0, & \text{if } k \leq 0. \end{cases}$$

The Cartesian product of $M_1 = (U_1, m_1)$ and $M_2 = (U_2, m_2)$ is defined as $M_1 \times M_2 = (U_1 \times U_2, m)$, where $m(x, y) = m_1(x)m_2(y)$ for all $x \in U_1$ and $y \in U_2$.

Relations

In mathematics, a subset of $U_1 \times U_2 \times \dots \times U_k$ is called a k -ary relation among the sets $U_1, U_2, \dots, U_k$, and a subset of U^k is called a k -ary relation on U . We are particularly interested in binary relations. According to the definition, a binary relation on U is a set of pairs of elements in U . A pair (x, y) in a binary relation means that x is related to y by the relation. We use lowercase Greek letters, α, β, ρ , etc., to denote relations. The basic operations on binary relations include the three aforementioned set-theoretic operations – intersection, union, and difference, as well as converse and composition. The converse of a binary relation $\alpha \subseteq U_1 \times U_2$ is $\alpha^- \subseteq U_2 \times U_1$ such that

$$\alpha^- = \{(x, y) | (y, x) \in \alpha\}.$$

The composition of two binary relations $\alpha \subseteq U_1 \times U_2$ and $\beta \subseteq U_2 \times U_3$ is $\alpha\beta \subseteq U_1 \times U_3$ such that

$$\alpha\beta = \{(x, y) | \exists z \in U, (x, z) \in \alpha \wedge (z, y) \in \beta\}.$$

A binary relation $\rho \subseteq U^2$ is said to be reflexive if $(x, x) \in \rho$ for all $x \in U$; ρ is symmetric if for all $x, y \in U$, $(x, y) \in \rho$ implies $(y, x) \in \rho$; ρ is anti-symmetric if for all $x, y \in U$, $(x, y) \in \rho$ and $(y, x) \in \rho$ imply $x = y$; and ρ is transitive if for all $x, y, z \in U$, $(x, y) \in \rho$ and $(y, z) \in \rho$ imply $(x, z) \in \rho$. A binary relation that is reflexive, anti-symmetric, and transitive is called a partial order. Let $\leq$ be a partial order on the universe U and $X \subseteq U$; then, $u \in U$ is called an upper bound (resp. lower bound) of X if for all $x \in X$, $x \leq u$ (resp. $u \leq x$). The least upper bound (resp. the greatest lower bound) of X is the upper bound (resp. lower bound) u of X such that for any upper bounds (resp. lower bounds) u' of X , $u \leq u'$ (resp. $u' \leq u$). The least upper bound (resp. the greatest lower bound) of X is called its supremum (resp. infimum) and is denoted by $\sup(X)$ (resp. $\inf(X)$). For any X (even though X is finite), $\sup(X)$ and $\inf(X)$ do not necessarily exist. If for all $x, y \in U$, $\sup(\{x, y\})$, and $\inf(\{x, y\})$ exist, then $\sup(\{x, y\})$ is called the join of x and y and denoted by $x \sqcup y$, while $\inf(\{x, y\})$ is called the meet of x and y and denoted by $x \sqcap y$. The structure $(U, \sqcap, \sqcup)$ is called a lattice. A subset $X \subseteq U$ is a sublattice of U if for any $x, y \in X$, $x \sqcap y$, and $x \sqcup y \in X$.

A binary relation that is reflexive, symmetric, and transitive is called an equivalence relation, which yields a partition of the universe. A partition of U is a class of subsets of U , $\{X_i | i \in I\}$ for some index set I , where $\cup_{i \in I} X_i = U$ and $X_i \cap X_j = \emptyset$ for any $i \neq j$. Let ρ be an equivalence relation on U , and let $x \in U$; then, the equivalence class containing x is defined as

$$[x]_\rho = \{y \in U | (x, y) \in \rho\}.$$

Obviously, $\{[x]_\rho | x \in U\}$ forms a partition of the universe. For any $X \subseteq U$, we denote $[X]_\rho$ by the set $\{[x]_\rho | x \in X\}$ and $\llbracket X \rrbracket_\rho$ by the multiset $([U]_\rho, m)$ such that $m([x]_\rho) = |[x]_\rho \cap X|$ for all $x \in U$.

Equivalence relations on a fixed domain are partially ordered by set inclusion. Let ρ_1 and ρ_2 be two equivalence relations and define $\rho_1 \leq \rho_2$ iff $\rho_1 \subseteq \rho_2$; then, $\leq$ is a partial order on the set of all equivalence relations on a fixed domain. We say that ρ_1 is finer than ρ_2 or ρ_2 is coarser than ρ_1 if $\rho_1 \leq \rho_2$. Given a binary relation α , the composition of α with itself k times is denoted by α^k , and the transitive closure of α is defined as $\alpha^\infty = \cup_{k \geq 1} \alpha^k$. Let ρ_1 and ρ_2 be two equivalence relations; then it can be shown that $\rho_1 \cap \rho_2$ and $(\rho_1 \cup \rho_2)^\infty$ are, respectively, the meet and the join of ρ_1 and ρ_2 .

Rough Set Theory: Functional Granulation

The basic construct of rough set theory is the approximation space, which is a pair (U, ρ) , where U is a finite set of objects (the universe) and ρ is an equivalence relation on U . From the GrC perspective, the equivalence classes of ρ are information granules of the approximation space.

In rough set theory, a concept is represented as a subset of the universe. For any concept $X \subseteq U$, we can associate the following two subsets with X ,

$$\begin{aligned} \underline{\rho}X &= \{x \in U | [x]_\rho \subseteq X\} \\ \overline{\rho}X &= \{x \in U | [x]_\rho \cap X \neq \emptyset\}, \end{aligned}$$

where $\underline{\rho}X$ and $\overline{\rho}X$ are called the ρ -lower and ρ -upper approximation of X , respectively. From a practical viewpoint, ρ can be considered as an indiscernibility relation; thus, for a given concept X , we can only know that X contains all the elements in $\underline{\rho}X$ and does not contain any element outside $\overline{\rho}X$. In other words, the concept X cannot be defined exactly in the approximation space.

The pair $(\underline{\rho}X, \overline{\rho}X)$ is considered a rough approximation of X , and any such pair is called a rough set. A concept X is called an exact concept in the approximation space if $\underline{\rho}X = X = \overline{\rho}X$; otherwise, it is called a rough concept.

Rough set theory is very useful in the analysis of data tables. A data table is a pair $S = (U, F)$, where U is a nonempty, finite set (the universe) and F is a nonempty, finite set of primitive attributes. Every $f \in F$ is a total function $f: U \rightarrow V_f$, where V_f denotes the possible values of f . If $E = \{f_1, \dots, f_n\} \subseteq F$, then $V_E = V_{f_1} \times \dots \times V_{f_n}$. Thus, E is also considered a function from U to V_E . An equivalence relation $IND(E)$ is associated with every subset of attributes $E \subseteq F$ and defined by

$$(x, y) \in IND(E) \Leftrightarrow f(x) = f(y) \forall f \in E.$$

$IND(E)$ is called an indiscernibility relation. We write $IND(f)$ instead of $IND(\{f\})$ for all $f \in F$. It is easy to show that $IND(E) = \bigcap_{f \in E} IND(f)$. Since $IND(E)$ is an equivalence relation, $(U, IND(E))$ is an approximation space, and we can define the $IND(E)$ -lower and $IND(E)$ -upper approximation of X for any $X \subseteq U$. We write $\underline{E}X$ and $\overline{E}X$ instead of $IND(E)X$ or $\overline{IND(E)}X$. Obviously, exact concepts in $(U, IND(E))$ are those definable using only attributes in E . In effect, the indiscernibility relation $IND(E)$ granulates (or partitions) the universe according to the functions in E ; thus, it results in a functional granulation of the universe.

The definitions are used in the analysis of dependency between attributes in a data table. Let us say that attribute E_2 depends on E_1 , denoted by $E_1 \Rightarrow E_2$, iff $IND(E_1) \subseteq IND(E_2)$, i.e., any two objects in U with the same attribute values in E_1 will also have the same attribute values in E_2 . It is easy to show that $E_1 \Rightarrow E_2$ iff $\underline{E_1}X = X$ for all X that are equivalence classes of $IND(E_2)$. Given an attribute $f \in F$, if $F - \{f\} \Rightarrow \{f\}$, then f can be eliminated from F without influencing the definability of F . In other words, by using the set of attributes, f is redundant for classification purposes. By repeated elimination of redundant

attributes, we can eventually obtain an irreducible set of attributes with the same classification capability as the original set of attributes. Such an irreducible set of attributes is called a reduct in rough set theory. Note that the definition of dependency is relative to a data table, so it does not necessarily mean that an intrinsic connection exists between the inter-dependent attributes.

Positional Analysis: Relational Granulation

Social networks are defined by actors and relations (or nodes and edges in terms of graph theory) (Hanneman and Riddle 2005). A social network is generally defined as a relational structure $\mathfrak{N} = (A, (\alpha_i)_{i \in I})$, where A is the set of actors in the network, I is an index set, and for each $i \in I$, $\alpha_i \subseteq A^{k_i}$ is a k_i -ary relation on the domain A . If $k_i = 1$, then α_i is also called an attribute. In practice, most SNA literature considers a simplified version of social networks with only binary relations. For ease of presentation, we focus on a social network with only unary and/or binary relations. Thus, the social network considered in this article is a structure $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$, where A is a finite set of actors, $P_i \subseteq A$ for all $i \in I$, and $\alpha_j \subseteq A \times A$ for all $j \in J$. In terms of graph theory, $\mathfrak{N}$ is a labeled graph, where A is a set of nodes labeled with subsets of I , and each α_j denotes a set of (labeled) edges. For each $a \in A$, the out-neighborhood and in-neighborhood of a with respect to a binary relation α , denoted, respectively, by $N_\alpha^+(a)$ and $N_\alpha^-(a)$, are defined as follows:

$$\begin{aligned} N_\alpha^+(a) &= \{b \in A \mid (a, b) \in \alpha\}, \\ N_\alpha^-(a) &= \{b \in A \mid (b, a) \in \alpha\}. \end{aligned}$$

If ρ is an equivalence relation on A and a is an actor, the ρ -equivalence class of a is equal to its neighborhood, i.e., $[a]_\rho = N_\rho^+(a) = N_\rho^-(a)$. Note that the latter equality holds because of the symmetry of ρ .

Several equivalence relations have been proposed for exploring the similarity between actors' roles. The simplest definition of positional equivalence is the concept of structural equivalence proposed in Lorrain and White (1971), which states that two actors are positionally equivalent if they are related to the same individuals.

Definition 1 Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network and ρ be an equivalence relation on A ; then ρ is a (strong) structural equivalence with respect to $\mathfrak{N}$ if $(a, b) \in \rho$ implies:

1. $a \in P_i$ iff $b \in P_i$ for all $i \in I$
2. $N_{\alpha_j}^+(a) = N_{\alpha_j}^+(b)$ and $N_{\alpha_j}^-(a) = N_{\alpha_j}^-(b)$ for all $j \in J$.

By the definition, there may exist more than one structural equivalence for a given network. However, it has been shown that there always exists a maximum (i.e., coarsest) structural equivalence for a network (Lerner 2005). In fact, the set of all structural equivalences form a sublattice of all equivalence relations on A . Indeed, if ρ_1 and ρ_2 are structural equivalences with respect to $\mathfrak{N}$, so too are $\rho_1 \sqcap \rho_2$ and $\rho_1 \sqcup \rho_2$. Since the set of all equivalence relations on a finite domain is finite, the join of all structural equivalences is the coarsest structural equivalence with respect to a network. Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network and ρ be the coarsest structural equivalence with respect to $\mathfrak{N}$; then, two actors x and $y \in A$ are said to be structurally equivalent, denoted by $x \cong_{\mathfrak{N}} y$, if $(x, y) \in \rho$.

Although structural equivalence is conceptually simple, it is sometimes restrictive. Regular equivalence relaxes the requirement that equivalent actors must be connected with identical actors and suggests that actors occupy the same position if they are connected to positionally equivalent actors. Regular equivalence has been studied extensively (Borgatti and Everett 1989; Boyd and Everett 1999; Everett and Borgatti 1994; Lerner 2005; White and Reitz 1983), and there are several alternative definitions of the concept. We present two here. The first is based on the

characterization given by Boyd and Everett (1999), which states that an equivalence relation ρ is a *regular equivalence* with respect to a binary relation α if it commutes with α , i.e.,

$$\alpha\rho = \rho\alpha.$$

By this definition, if ρ is a regular equivalence with respect to α and $(a, b) \in \rho$, then for each $c \in N_{\alpha}^+(a)$ (resp. $N_{\alpha}^-(a)$), there exists $c' \in N_{\alpha}^+(b)$ (resp. $N_{\alpha}^-(b)$) such that $(c, c') \in \rho$. The property naturally leads to an alternative definition of regular equivalence (Lerner 2005), which states that an equivalence relation ρ is a *regular equivalence* with respect to a binary relation α if for $a, b \in A$,

$$\begin{aligned} (a, b) \in \rho &\Rightarrow [N_{\alpha}^+(a)]_{\rho} \\ &= [N_{\alpha}^+(b)]_{\rho} \text{ and } [N_{\alpha}^-(a)]_{\rho} \\ &= [N_{\alpha}^-(b)]_{\rho}. \end{aligned}$$

According to this definition, if a and b are regularly equivalent, then they are connected to equivalent neighborhoods. Obviously, the above definitions are equivalent. Thus, we have the following definition.

Definition 2 Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network and ρ be an equivalence relation on A ; then ρ is a *regular equivalence* with respect to $\mathfrak{N}$ if:

1. $(a, b) \in \rho$ implies $a \in P_i$ iff $b \in P_i$ for all $i \in I$
2. ρ is a regular equivalence with respect to α_j for all $j \in J$

Analogous to the case of structural equivalences, the join of two regular equivalences is still a regular equivalence, even though their meet is not necessarily a regular equivalence. Consequently, we can find the coarsest regular equivalence of a network. Then, two actors, x and y , are regularly equivalent, denoted by

$x \cong_r y$, if (x, y) is in the coarsest regular equivalence of the network.

For regular equivalence, only the occurrence or non-occurrence of a position in the neighborhood of an actor matters. However, the number of occurrences is sometimes an important factor in positional analysis. In such cases, a number restriction can be added to the definition of regular equivalences. An equivalence relation ρ is an *exact equivalence* with respect to a binary relation α if for $a, b \in A$,

$$(a, b) \in \rho \Rightarrow N_{\alpha}^{+}(a)_{\rho} = N_{\alpha}^{+}(b)_{\rho} \text{ and } N_{\alpha}^{-}(a)_{\rho} = N_{\alpha}^{-}(b)_{\rho}.$$

By replacing the set equality in the definition with the multiset equality, the number of equivalent neighbors must be the same for two actors to be considered position-equivalent.

Definition 3 Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network and ρ be an equivalence relation on A ; then, ρ is an exact equivalence with respect to $\mathfrak{N}$ if:

1. $(a, b) \in \rho$ implies $a \in P_i$ iff $b \in P_i$ for all $i \in I$
2. ρ is an exact equivalence with respect to α_j for all $j \in J$

The set of all exact equivalences also forms a lattice, but it is not a sublattice of the set of all equivalence relations (Everett and Borgatti 1994). Consequently, the coarsest exact equivalence for a network exists, and we can define two actors x and y as exactly equivalent, denoted by $x \cong_e y$, if they belong to the coarsest exact equivalence. The partition produced by an exact equivalence is called an equitable partition or a divisor of a graph (Lerner 2005).

From the above definition, it is clear that all the attributes in a social network are binary. In other words, for each attribute P_i , an actor either satisfies P_i (has the value 1) or falsifies P_i (has the value 0). When a social

network $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ has the property $J = \emptyset$, it is in fact a kind of data table. In this case, all three kinds of positional equivalence are reduced to the indiscernibility relation. This means that, if information about the actors' relationship is not available, social positions are fully determined by the attributes of the actors. Thus, without relational information, functional granulation is sufficient for positional analysis. However, if relational information is available, two actors can be further differentiated by their relationship with other actors, even though all their attributes have the same values. In this sense, the definition of positional equivalence is based on the process of relational granulation.

Modal Logic: The Bridge

Modal logics were originally developed as formalizations for reasoning about modalities. The formal study started with alethic modal logics (the logics of necessity and possibility). The modalities $\Box$ (for necessity) and $\Diamond$ (for possibility) have now become standard notations in modal logic literature, and many variants of modal logic have been proposed to deal with different classes of modalities, for example, temporal logic for tense operators, deontic logic for obligation and permission, dynamic logic for actions, and epistemic logic for beliefs and knowledge. The development of these logics was motivated by philosophical enquiry as well as by technical applications in computer science, artificial intelligence, and economic game theory.

While modal logics were initially presented in the form of reasoning systems, the invention of relational semantics has been influential in the continuing development of the field (Kripke 1959, 1963). The semantics, proposed independently by J. Hintikka, S. Kanger, and S. Kripke, and now known as Kripke semantics, show that modal logics are in fact logics for reasoning about relational structures. From this semantic perspective, standard modal logics can be regarded as

fragments of first- or second-order predicate logics, whereby the necessity and possibility modalities correspond to universal and existential quantifiers, respectively. In spite of this correspondence, quantification in modal logic tends to be bounded in some way to worlds that are “relevant to” or “accessible from” the current one. Consequently, although a number of properties of modal logics follow immediately from those of their classical quantificational counterparts, the modal operators typically have less expressive power than full quantification. This results in many interesting properties not available in classical predicate logic. One of the most striking results is that semantic invariances between models are actually various forms of *bisimulation*, which preserve the local properties of worlds and their transition patterns (Blackburn et al. 2001). Interestingly, it has been shown that bisimulation in Kripke models corresponds exactly to regular equivalence in social networks (Marx and Masuch 2003). The implication of the results is that position-equivalent actors can be characterized by an appropriate set of modal formulas. As each modal formula can be regarded as a binary attribute, from the GrC perspective, such characterization transforms relational granulation into functional granulation.

In this section, we present three modal logics – a multi-modal logic, a graded modal logic, and a hybrid logic. The presentation of a modal logic, as with any other formal logic, requires the specification of its syntax and semantics. The syntactic aspect includes the language of the logic – its alphabet and formation rules for wffs, as well as a deductive system for logical reasoning. For the purpose of this article, we need only present the language part. For the semantic aspect, we have to define the models within which the wffs can be interpreted, as well as the satisfaction of a wff in a model.

Propositional Multi-modal Logic

We start with propositional modal logic (PML) (Chellas 1980). The alphabet of PML consists of a set of propositional symbols, PV , and the

logical symbols $\neg$ (negation), $\wedge$ (and), $\vee$ (or), $\supset$ (material implication), and $\Diamond$ (possibility); $\neg$, $\wedge$, $\vee$, and $\supset$ are called Boolean connectives, whereas $\Diamond$ is the only primitive modality of PML. The set of wffs of PML is the smallest set containing PV that satisfies the following conditions:

- If φ is a wff, then $\neg\varphi$ and $\Diamond\varphi$ are wffs.
- If φ and ψ are wffs, then $\varphi \wedge \psi$, $\varphi \vee \psi$, and $\varphi \supset \psi$ are wffs.

As usual, we abbreviate $(\varphi \supset \psi) \wedge (\psi \supset \varphi)$ as $\varphi \equiv \psi$; $\neg \Diamond \neg \varphi$ as $\Box\varphi$; any tautology $p \vee \neg p$ as $\top$; and any contradiction $p \wedge \neg p$ as $\perp$. A Kripke model for PML is a triple $\mathfrak{M} = (W, \alpha, V)$, where W is a set of possible worlds; α is a binary relation on W , called an accessibility relation; and $V: W \times PV \rightarrow \{0, 1\}$ is a truth assignment for evaluating the truth value of each propositional symbol in each world. The satisfaction of a wff φ in a world w of the model $\mathfrak{M}$, denoted by $\mathfrak{M}, w \models \varphi$, is defined by the following clauses:

1. $\mathfrak{M}, w \models p$ if $V(w, p) = 1$ for each $p \in PV$.
2. $\mathfrak{M}, w \models \neg\varphi$ iff $\mathfrak{M}, w \not\models \varphi$.
3. $\mathfrak{M}, w \models \varphi \wedge \psi$ iff $\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$.
4. $\mathfrak{M}, w \models \varphi \vee \psi$ iff $\mathfrak{M}, w \models \varphi$ or $\mathfrak{M}, w \models \psi$.
5. $\mathfrak{M}, w \models \varphi \supset \psi$ iff $\mathfrak{M}, w \not\models \varphi$ or $\mathfrak{M}, w \models \psi$.
6. $\mathfrak{M}, w \models \Diamond\varphi$ iff there exists $(w, u) \in \alpha$ such that $\mathfrak{M}, u \models \varphi$.

Propositional multi-modal logic (PMML) is an extension of PML that allows more than one modality (Horrocks et al. 2007). PMML is particularly suitable for reasoning about relational structures that contain a number of binary relations, e.g., social networks. The alphabet of PMML consists of a set of propositional symbols PV , a set of relational symbols REL , the Boolean connectives, the relational converse symbol $\bar{}$, and the modality-forming symbol $\langle \rangle$. The set of wffs of PMML is the smallest set containing PV that satisfies the following conditions:

- If φ is a wff, then $\neg\varphi$ is a wff.
- If φ is a wff and r is a relational symbol, then $\langle r \rangle\varphi$ and $\langle r^- \rangle\varphi$ are wffs.
- If φ and ψ are wffs, then $\varphi \wedge \psi$, $\varphi \vee \psi$, and $\varphi \supset \psi$ are wffs.

We abbreviate $\neg\langle r \rangle \neg\varphi$ and $\neg\langle r^- \rangle \neg\varphi$ as $[r]\varphi$ and $[r^-]\varphi$, respectively. A Kripke model for PMML is $\mathfrak{M} = (W, (\alpha_r)_{r \in REL}, V)$, where W and V are the same as above, and for each $r \in REL$, α_r is a binary relation on W . The satisfaction of PMML wffs is defined in the same way as that of PML wffs, except that clause 6 is replaced by:

6. $\mathfrak{M}, w \models \langle r \rangle\varphi$ iff there exists $(w, u) \in \alpha_r$ such that $\mathfrak{M}, u \models \varphi$.
7. $\mathfrak{M}, w \models \langle r^- \rangle\varphi$ iff there exists $(u, w) \in \alpha_r$ such that $\mathfrak{M}, u \models \varphi$.

Note that we need the converse modalities $\langle r^- \rangle$ because all positional equivalences are defined with respect to their in-neighborhoods and out-neighborhoods. The modalities $\langle r \rangle$ allow us to access the out-neighborhood of a world, whereas $\langle r^- \rangle$ is needed to access the in-neighborhood.

Graded Modal Logic

Graded modal logic (GML) is a generalization of PMML that allows us to consider the number of worlds accessible from a given world (van der Hoek 1992). The alphabet of GML is the same as that of PMML; however, the second formation rule of wffs is replaced by the following rule:

- If φ is a wff and r is a relational symbol, then $\langle r \rangle_n\varphi$ and $\langle r^- \rangle_n\varphi$ are wffs for each natural number n .

We also write $[r]_n\varphi$ and $[r^-]_n\varphi$ as the abbreviations of $\neg\langle r \rangle_n \neg\varphi$ and $\neg\langle r^- \rangle_n \neg\varphi$, respectively.

The Kripke models for GML are the same as those for PMML, but clauses 6 and 7 for the conditions of satisfaction of PMML wffs are modified as follows:

6. $\mathfrak{M}, w \models \langle r \rangle_n\varphi$ iff $|\{u \in W \mid (w, u) \in \alpha_r\}| > n$.

7. $\mathfrak{M}, w \models \langle r^- \rangle_n\varphi$ iff $|\{u \in W \mid (u, w) \in \alpha_r \text{ and } \mathfrak{M}, u \models \varphi\}| > n$.

According to the semantics, the modalities $\langle r \rangle_0$ and $\langle r^- \rangle_0$ in GML are, respectively, equivalent to $\langle r \rangle$ and $\langle r^- \rangle$ in PMML. Intuitively, $\mathfrak{M}, w \models \langle r \rangle_n\varphi$ means that φ is true in more than n worlds that are r -accessible from w , whereas $\mathfrak{M}, w \models [r]_n\varphi$ means that φ is false in no more than n worlds that are r -accessible from w . Thus, we abbreviate $\langle r \rangle_{n-1}\varphi \wedge [r]_n \neg\varphi$ (resp. $\langle r^- \rangle_{n-1}\varphi \wedge [r^-]_n \neg\varphi$) as $(r)_n\varphi$ (resp. $(r^-)_n\varphi$) to denote that there are exactly n r -accessible (resp. r^- -accessible) worlds that satisfy φ .

Hybrid Logic

Hybrid logics (HL) are extensions of modal logics that include symbols to name possible worlds in models. The development of HL dates back to A. Prior's 1951 work, which had a philosophical foundation (Prior 1967, 1968) (see Areces and ten Cate (2007) for a brief overview). The main idea is to use a special kind of atomic formula, called a nominal, to refer to possible worlds in models. The idea was adopted by the Sofia school in the late 1980s and early 1990s (Gargov and Goranko 1993; Passy and Tinchev 1985, 1991). Subsequently, Blackburn and Seligman, who have been influential in the recent development of HL, adopted it in the late 1990s (Blackburn 2000; Blackburn and Seligman 1995).

Here, we only need the simplest hybrid logic (SHL). The alphabet of SHL is the same as that of PMML plus a set of nominals NOM that is disjoint with the set PV . The set of wffs of SHL is the smallest set containing $PV \cup NOM$ that satisfies the formation rules of PMML. The key semantic idea of HL is that each nominal must be true in exactly one possible world in any model. In other words, a nominal names a world by being true in that world and nowhere else. The idea can be easily formalized by extending PMML models with a component that assigns a nominal to a possible world. A hybrid model is then $\mathfrak{M} = (W, (\alpha_r)_{r \in REL}, V, f)$, where $(W, (\alpha_r)_{r \in REL}, V)$ is a PMML model and $f: NOM \rightarrow W$ is the mapping that assigns a nominal to a possible world. The

satisfaction of SHL wffs is defined by clauses 1–7 for PMML, with the addition of the following clause for nominals:

8. $\mathfrak{M}, w \mid = i \text{ iff } (i) = w \text{ for each } i \in \text{NOM}$.

From Relational Granulation to Functional Granulation

In this section, we explain how to transform positional equivalences from relational granulation to functional granulation. Given a social network $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$, the steps are as follows:

1. For each kind of equivalence, find a corresponding modal language based on $\mathfrak{N}$. The three kinds of positional equivalence discussed earlier correspond to the following logical languages:
 - (a) Structural equivalence: SHL
 - (b) Regular equivalence: PMML
 - (c) Exact equivalence: GML
2. Consider $\mathfrak{N}$ as a model in the corresponding logic by a simple transformation. The set of actors is the set of possible worlds in the resultant model.
3. Find a sublanguage $\mathcal{L}$ of the corresponding logic. Prove that for any two actors x and y , $x \cong y$ iff x and y satisfy the same set of wffs in $\mathcal{L}$, where $\cong$ is $\cong_g, \cong_r, \cong_e$.

Structural Equivalences and Hybrid Logics

We use SHL to characterize structural equivalences. Given a social network $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$, we define an SHL language with the following alphabet:

1. $PV = \{p_i \mid i \in I\}$.
2. $REL = J$.
3. $NOM = A$.

The social network $\mathfrak{N}$ is transformed into a hybrid model $\mathfrak{M}_{\mathfrak{N}} = (A, (\alpha_j)_{j \in J}, V, id)$ where

V is defined by $V(x, p_i) = 1$ iff $x \in P_i$ for all $x \in A$ and $i \in I$, and id is the identity function on A . Let us now define a sublanguage $\mathcal{H}_{\mathfrak{N}}$ of the SHL as the set of wffs $PV \cup \{\langle j \rangle a, \langle j^- \rangle a \mid j \in REL, a \in NOM\}$. We say that two actors, x and y , are $\mathcal{H}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$ if for all $\varphi \in \mathcal{H}_{\mathfrak{N}}$, $(\mathfrak{M}_{\mathfrak{N}}, x \mid = \varphi$ iff $\mathfrak{M}_{\mathfrak{N}}, y \mid = \varphi)$. Then, we have the following characterization theorem:

Theorem 1 *Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network; then, for all $x, y \in A$, $x \cong_s y$ in $\mathfrak{N}$ iff x and y are $\mathcal{H}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$.*

Proof By the satisfaction condition, $(\mathfrak{M}_{\mathfrak{N}}, x \mid = \langle j \rangle a$ iff $(x, a) \in \alpha_j$. Thus, x and y are $\mathcal{H}_{\mathfrak{N}}$ -equivalent iff the following two conditions are satisfied:

1. $x \in P_i$ iff $y \in P_i$ for all $i \in I$ (as $PV \subseteq \mathcal{H}_{\mathfrak{N}}$).
2. For all $j \in J$ and $a \in A$, $(x, a) \in \alpha_j$ iff $(y, a) \in \alpha_j$ and $(a, x) \in \alpha_j$ iff $(a, y) \in \alpha_j$ (since both $\langle j \rangle a, \langle j^- \rangle a$ and $\langle j \rangle a, \langle j^- \rangle a$ are in $\mathcal{H}_{\mathfrak{N}}$).

These two conditions are exactly the same as those in Definition 1. Since $x \cong_s y$ iff (x, y) is in the coarsest structural equivalence, $x \cong_s y$ iff these two conditions are satisfied. ■

Regular Equivalences and Multi-modal Logics

The first logical characterization of positional equivalences is that of regular equivalences by dynamic logic – a particular type of PMML. This is shown in Marx and Masuch (2003) by the connection between regular equivalence and the well-known notion of bisimulation in modal logics. We present the result here by following the general procedure of the transformation from relational granulation to functional granulation.

Given a social network $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$, we define a PMML language with the following alphabet:

1. $PV = \{p_i \mid i \in I\}$.
2. $REL = J$.

The social network $\mathfrak{N}$ is transformed into a Kripke model $\mathfrak{M}_{\mathfrak{N}} = (A, (\alpha_j)_{j \in J}, V)$, where V is defined by $V(x, p_i) = 1$ iff $x \in P_i$ for all $x \in A$ and $i \in I$. Let $\mathcal{L}_{\mathfrak{N}}$ be the set of wffs of the PMML and $\mathcal{L}_{\mathfrak{N}}$ -equivalence be defined the same as $\mathcal{H}_{\mathfrak{N}}$ -equivalence.

Theorem 2 (Marx and Masuch 2003) *Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network; then, for all $x, y \in A$, $x \cong_{\rho} y$ in $\mathfrak{N}$ iff x and y are $\mathcal{L}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$.*

Proof It is shown in Marx and Masuch (2003) that regular equivalences are exactly bisimulations, i.e., $x \cong_{\rho} y$ iff x and y are bisimilar. By Theorems 2.20 and 2.24 in Blackburn et al. (2001), x and y are bisimilar iff they are $\mathcal{L}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$. ■

Exact Equivalences and Graded Modal Logics

Following the procedure used in the previous subsection, we define $\mathcal{G}_{\mathfrak{N}}$ as the set of wffs of the GML based on the given network $\mathfrak{N}$. Then, we have the following theorem:

Theorem 3 *Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network; then, for all $x, y \in A$, $x \cong_{\rho} y$ in $\mathfrak{N}$ iff x and y are $\mathcal{G}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$.*

Proof 1. $(\Rightarrow)$: Assume that $x \cong_{\rho} y$; then, there exists an exact equivalence ρ on $\mathfrak{N}$ such that $(x, y) \in \rho$. We show that $\mathfrak{M}_{\mathfrak{N}}, x \models \varphi$ implies $\mathfrak{M}_{\mathfrak{N}}, y \models \varphi$ by induction on the complexity of φ (the proof of the converse direction is exactly the same). The case where $\varphi \in PV$ follows from condition 1 of Definition 3. The cases for Boolean connectives are straightforward by the induction hypothesis. For formulas of the form $\langle j \rangle_n \psi$ or $\langle j^- \rangle_n \psi$, we prove the case for $\langle j \rangle_n \psi$ by using the condition $N_{\alpha_j}^+(x)_{\rho} = N_{\alpha_j}^+(y)_{\rho}$. The case for $\langle j^- \rangle_n \psi$ is proved similarly by applying $N_{\alpha_j}^-(x)_{\rho} = N_{\alpha_j}^-(y)_{\rho}$. Now, if $\mathfrak{M}_{\mathfrak{N}}, x \models \langle j \rangle_n \psi$, then there exist more than n worlds in $N_{\alpha_j}^+(x)$ that satisfy ψ . Assume

$u_0, u_1, \dots, u_n$ are among these worlds. Since $N_{\alpha_j}^+(x)_{\rho} = N_{\alpha_j}^+(y)_{\rho}$, there exist $u'_0, u'_1, \dots, u'_n \in N_{\alpha_j}^+(y)$ such that $(u_i, u'_i) \in \rho$ for $0 \leq i \leq n$. By the induction hypothesis, $\mathfrak{M}_{\mathfrak{N}}, u'_i \models \psi$ for $0 \leq i \leq n$; thus, $\mathfrak{M}_{\mathfrak{N}}, y \models \langle j \rangle_n \psi$.

2. $(\Leftarrow)$: Let us define an equivalence relation ρ on A by

$(x, y) \in \rho$ iff x and y are $\mathcal{G}_{\mathfrak{N}}$ -equivalent with respect to $\mathfrak{N}$.

We prove that ρ is an exact equivalence. The first condition of Definition 3 is straightforward. For the second condition, let us assume that $(x, y) \in \rho$, $N_{\alpha_j}^+(x)_{\rho} = ([A]_{\rho}, m_1)$, and $N_{\alpha_j}^+(y)_{\rho} = ([A]_{\rho}, m_2)$ for some $j \in J$. Assume also that $[A]_{\rho} = \{[z_i]_{\rho} \mid 1 \leq i \leq k\}$. First, we consider z_1 as an example. For each $2 \leq i \leq k$, since $[z_i]_{\rho}$ is an equivalence class distinct from $[z_1]_{\rho}$, there exists a $\mathcal{G}_{\mathfrak{N}}$ -wff ψ_i such that $\mathfrak{M}_{\mathfrak{N}}, z_i \models \psi_i$, but $\mathfrak{M}_{\mathfrak{N}}, z_1 \models \neg \psi_i$. If $m_1([z_1]_{\rho}) = n_1 > 0$, then there exist exactly n_1 α_j -successors of x that are $\mathcal{G}_{\mathfrak{N}}$ -equivalent to z_1 . Thus, $\mathfrak{M}_{\mathfrak{N}}, x \models (j)_{n_1}(\psi \wedge \bigwedge_{i=2}^k \neg \psi_i)$ for all ψ satisfied in z_1 . Since $(x, y) \in \rho$, it is that $\mathfrak{M}_{\mathfrak{N}}, y \models (j)_{n_1}(\psi \wedge \bigwedge_{i=2}^k \neg \psi_i)$ for all ψ satisfied in z_1 . This means there are exactly n_1 α_j -successors of y that satisfy $\bigwedge_{i=2}^k \neg \psi_i$ (by letting $\psi = \top$). Thus, these n_1 α_j -successors of y must satisfy all ψ satisfied in z_1 and be $\mathcal{G}_{\mathfrak{N}}$ -equivalent to z_1 . Since no other α_j -successors of y can be $\mathcal{G}_{\mathfrak{N}}$ -equivalent to z_1 (otherwise there would be more than n_1 α_j -successors of y satisfying $\bigwedge_{i=2}^k \neg \psi_i$), $m_2([z_1]_{\rho}) = n_1$. In the same way, we can prove that if $m_2([z_1]_{\rho}) > 0$, then $m_1([z_1]_{\rho}) > 0$. Thus, $m_1([z_1]_{\rho}) = m_2([z_1]_{\rho})$. Similarly, we can prove $m_1([z_i]_{\rho}) = m_2([z_i]_{\rho})$ for $2 \leq i \leq k$. Consequently, we have $m_1 = m_2$ and $N_{\alpha_j}^+(x)_{\rho} = N_{\alpha_j}^+(y)_{\rho}$. The proof of $N_{\alpha_j}^-(x)_{\rho} = N_{\alpha_j}^-(y)_{\rho}$ is analogous, except that the modality $(j)_{n_1}$ is replaced by $(j^-)_{n_1}$.

Observational Equivalences

We have presented the logical characterizations of three kinds of positional equivalence. The theoretical implication of these characterizations is

that positional equivalences based on relational granulation can be transformed into indiscernibility relations based on a set of binary attributes. Let $\mathfrak{N} = (A, (P_i)_{i \in I}, (\alpha_j)_{j \in J})$ be a social network and let F denote $\mathcal{H}_{\mathfrak{N}}$, $\mathcal{L}_{\mathfrak{N}}$, or $\mathcal{G}_{\mathfrak{N}}$. Then, (A, F) is a (generalized) data table such that $IND(F)$ is the corresponding positional equivalence. However, from a practical viewpoint, since both $\mathcal{L}_{\mathfrak{N}}$ and $\mathcal{G}_{\mathfrak{N}}$ are infinite sets, it is generally difficult (if not impossible) to decide $IND(F)$ based on such a data table representation. This motivates us to define a new kind of positional equivalence based on observations.

Let $\mathfrak{N}$ be a social network and F be a *finite* set of wffs in an appropriate logic language. Then, two actors, x and y , are *F-observationally equivalent*, denoted by $x \cong_F y$, if for any wff $\varphi \in F$, x satisfies φ iff y satisfies φ . The wffs in F denote the observable properties of the actors in the network. Due to the limited capability of the observer, the set of observable properties is usually finite. Depending on the observable properties chosen, we can focus on different parts of the social network. For example, positional analysis for a genealogical study and a political investigation may have to consider different relationships between actors. As shown by the above results, the choice of an appropriate logic language is also crucial, since it may result in different notions of positional equivalence.

A recent approach to structural equivalence that conforms with our definition of observational equivalence is reported in Pizarro (2007). In that approach, a set of institutions E is given and each actor may belong to several institutions. Two actors are structurally equivalent if they belong to the same set of institutions. To avoid confusion with the structural equivalence defined in section “Positional Analysis: Relational Granulation,” we call the equivalence relation defined in Pizarro (2007) *institutional equivalence*. Note that institutional equivalence is in fact the same as structural equivalence defined in section “Positional Analysis: Relational Granulation,” if we consider the universe as the union of A and E and represent the network as a bipartite graph. A place (social position) is then defined as an equivalence class of

actors (or, equivalently, the set of institutions that those actors belong to). Here, our set of observation sentences corresponds to the set of institutions in Pizarro (2007). While institutions are primitive entities, our observation sentences are formed from the social relations and attributes of actors. In fact, a wff in F can also be regarded as an intensional definition of an institution. In this sense, observational equivalence is a kind of intensionally institutional equivalence.

Related Work

Since the publication of the first version of the article, there has been significant progress on the theory and applications of social network analysis by granular computing. Hence, in this updated version, we will briefly mention several important recent works in this area.

On the practical aspect, because of the close correspondence between modal logic and description logic (DL) (Baader and Nutt 2002; Nardi and Brachman 2002), a straightforward application of the GrC methodology to DL-based representation of social network is expectable. In particular, the notion of indiscernibility has been applied to privacy-preserving social network publication in Hsu et al. (2013, 2014). In addition, it is also shown that the GrC approach can be applied to concept mining from social networks (Fan 2013) and DL can provide a knowledge representation formalism for the results of data mining (Fan and Liao 2016).

On the theoretical aspect, the main development is the generalization to weighted social networks, which can represent not only qualitative relationships but also degrees of connection between actors (Barrat et al. 2004; Ćirić and Bogdanović 2010; Fan et al. 2007, 2008; Nair and Sarasamma 2007; Toivonen et al. 2007). The notion of regular equivalence is generalized to weighted social networks and connected with bisimulation in fuzzy modal logic in Fan and Liao (2013, 2014). This line of work is further elaborated in Ignjatović et al. (2015a, b) and strongly related to the solution of fuzzy relation equations from a computational viewpoint (Ćirić et al. 2012;

Ignjatović and Ćirić 2012; Ignjatović et al. 2013). From a purely logical perspective, the notion of bisimulation has also received considerable attention in the study of many-valued modal logic (Eleftheriou et al. 2012; Fan 2015; Marti and Metcalfe 2014; Wild et al. 2018).

Conclusion and Future Directions

This article starts with an introduction to rough set theory – the basic theory of functional granulation. We then review several notions of positional equivalence in social network analysis and consider them as relational granulation from the GrC perspective. By using modal logics as a bridge, we transform the relational granulation-based positional equivalences into functional granulation. However, the transformation is mainly theoretically interesting, since the set of functional attributes used to granulate a social network may be infinite. This practical consideration motivates us to propose the notion of observational equivalence. It conforms with an alternative definition of structural equivalence, which we call institutional equivalence to avoid confusion.

More generally, both observational equivalence and institutional equivalence can be analyzed in the framework of formal concept analysis (FCA) (Ganter and Wille 1998). FCA is a conceptual knowledge representation and data analysis method in which a *formal context* is defined as a triple (A, F, R) , where A and F are sets of objects and attributes, respectively; and $R \subseteq A \times F$ is a binary relation between A and F , called the incidence of the formal context. For $A' \subseteq A$ and $F' \subseteq F$, let us denote $R^-(A') = \{f \in F \mid \forall a \in A', (a, f) \in R\}$ and $R^+(F') = \{a \in A \mid \forall f \in F', (a, f) \in R\}$. Then, a formal concept is defined as a pair (A', F') such that $A' \subseteq A$, $F' \subseteq F$, $R^-(A') = F'$, and $R^+(F') = A'$. Here, we call A' the extent and F' the intent of the formal concept. The application of FCA to sociology has been proposed in Freeman and White (1993). In that application, the formal context (also called the Galois lattice) is used to model two-mode network data, where the set of actors is a mode of the data, the set of events is the other

mode of data, and the relation between the two modes is the participation relationship between the actors and events. When we regard A as the set of actors, F as the set of observation sentences (resp. institution), and R as the satisfaction (resp. membership) relation, a formal context can be constructed out of a social network. However, a social position (or place) defined by observational or institutional equivalences may not be the extent of a formal concept. Thus, it is tempting to define a social position alternatively as a formal concept of such formal contexts. Unlike other positional analysis techniques that define a social position as an equivalence class of some positional equivalence, the extents of different formal concepts may overlap. This means that social positions may not be disjoint in such a definition. The theoretical and practical implications of such positional analysis requires further research.

Another future research direction concerns the practical computation of positional equivalences. As mentioned earlier, when a social network is transformed into a data table (A, F) for positional analysis and F is infinite, it is difficult to decide if two actors are $IND(F)$ -equivalent by exhaustively comparing the satisfaction of all wffs in x and y . While observational equivalence provides a practical approximation of the computation, we would like to know if precise computation is possible. To answer this question, the inference of the *most specific concept* (msc) in description logics (Baader and Küsters 2006) may be helpful. Description logics are a family of knowledge representation formalisms that have strong connections with modal logics (Baader and Nutt 2002; Nardi and Brachman 2002). Indeed, the description languages $\mathcal{ALC}$ and $\mathcal{ALCN}$ correspond to the sublanguages of PMML and GML, respectively (Baader and Nutt 2002; Nardi and Brachman 2002). Given a network represented in a description language, the msc of an actor x , denoted by $msc(x)$, corresponds to a wff that satisfies x and implies any wffs that satisfy x . Thus, if $msc(x)$ and $msc(y)$ are logically equivalent, then x and y will satisfy the same set of wffs, i.e., $(x, y) \in IND(F)$. The problem for researchers is to fold logical languages that can characterize the aforementioned positional equivalences

adequately and whose msc problems can be solved effectively.

In summary, we have proposed a logical approach to positional analysis in social networks from a GrC perspective. The approach facilitates a connection between more complicated relational granulation and simpler functional granulation. While there remain problems to be solved in the research program, the approach seems promising for cross-fertilization between the fields of GrC and SNA.

Bibliography

- Areces C, ten Cate B (2007) Hybrid logics. In: Blackburn P, van Benthem J, Wolter F (eds) *Handbook of modal logic*. Elsevier, pp 821–868
- Baader F, Küsters R (2006) Nonstandard inferences in description logics: the story so far. In: Gabbay DM, Goncharov SS, Zakharyashev M (eds) *Mathematical problems from applied logic I: logics for XXIst century*, volume 4 of international mathematical series, pp 1–75. Springer
- Baader F, Nutt W (2002) Basic description logics. In: Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (eds) *Description logic handbook*, Cambridge University Press, pp 47–100
- Barrat A, Barthélémy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci* 101(11):3747–3752
- Blackburn P (2000) Representation, reasoning, and relational structures: a hybrid logic manifesto. *Log J IGPL* 8(3):339–625
- Blackburn P, de Rijke M, Venema Y (2001) *Modal logic*. Cambridge University Press
- Blackburn P, Seligman J (1995) Hybrid languages. *J Log Lang Inf* 4:251–272
- Borgatti SP, Everett MG (1989) The class of all regular equivalences: algebraic structure and computation. *Soc Netw* 11(1):65–88
- Boyd JP, Everett MG (1999) Relations, residuals, regular interiors, and relative regular equivalence. *Soc Netw* 21(2):147–165
- Chellas BF (1980) *Modal logic: an introduction*. Cambridge University Press
- Čirić M, Bogdanović S (2010) Fuzzy social network analysis. *Godišnjak Učiteljskog fakulteta u Vranju* 1:179–190
- Čirić M, Ignjatović J, Jančić I, Damjanović N (2012) Computation of the greatest simulations and bisimulations between fuzzy automata. *Fuzzy Sets Syst* 208:22–42
- Eleftheriou PE, Koutras CD, Nomikos C (2012) Notions of bisimulation for Heyting-valued modal languages. *J Log Comput* 22(2):213–235
- Everett MG, Borgatti SP (1994) Regular equivalences: general theory. *J Math Sociol* 18(1):29–52
- Fan TF (2013) Rough set analysis of relational structures. *Inf Sci* 221:230–244
- Fan TF (2015) Fuzzy bisimulation for Gödel modal logic. *IEEE Trans Fuzzy Syst* 23(6):2387–2396
- Fan TF, Liao CJ (2013) Many-valued modal logic and regular equivalences in weighted social networks. In: *Proceedings of the 12th European conference on symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU)*, vol 7958 of *Lecture notes in computer science*, pp 194–205. Springer
- Fan TF, Liao CJ (2014) Logical characterizations of regular equivalence in weighted social networks. *Artif Intell* 214:66–88
- Fan TF, Liao CJ (2016) Rough set-based concept mining from social networks. In: *Proceedings of the IEEE international conference on fuzzy systems (FUZZ-IEEE)*, pp 663–670
- Fan TF, Liao CJ, Lin TY (2007) Positional analysis in fuzzy social networks. In: *Proceedings of the 3rd IEEE international conference on granular computing*, pp 423–428
- Fan TF, Liao CJ, Lin TY (2008) A theoretical investigation of regular equivalences for fuzzy graphs. *Int J Approx Reason* 49(3):678–688
- Fan TF, Liao CJ, Liu DR, Tzeng GFL (2006) Granulation based on hybrid information systems. In: *Proceedings of the 2006 IEEE international conference on systems, man, and cybernetics*, pp 4768–4772
- Freeman LC, White DR (1993) Using Galois lattices to represent network data. *Sociol Methodol* 23:127–146
- Ganter B, Wille R (1998) *Formal concept analysis: mathematical foundations*. Springer
- Gargov G, Goranko V (1993) Modal logic with names. *J Philos Log* 22(6):607–636
- Hanneman RA, Riddle M (2005) *Introduction to social network methods*. University of California, Riverside
- Horrocks I, Hustadt U, Sattler U, Schmidt R (2007) Computational modal logic. In: Blackburn P, van Benthem J, Wolter F (eds) *Handbook of modal logic*. Elsevier, pp 181–245
- Hsu TS, Liao CJ, Wang DW (2013) Privacy-preserving social network publication based on positional indiscernibility. In: *Proceedings of the 7th international conference on scalable uncertainty management (SUM)*, volume 8078 of *lecture notes in computer science*, pp 311–324. Springer
- Hsu T s, Liao CJ, Wang DW (2014) A logical framework for privacy-preserving social network publication. *J Appl Log* 12(2):151–174
- Ignjatović J, Čirić M (2012) Weakly linear systems of fuzzy relation inequalities and their applications: a brief survey. *Filomat* 26(2):207–241
- Ignjatović J, Čirić M, Simović V (2013) Fuzzy relation equations and subsystems of fuzzy transition systems. *Knowl Based Syst* 38:48–61
- Ignjatović J, Čirić M, Stanković I (2015a) Bisimulations in fuzzy social network analysis. In: *Proceedings of the*

- 2015 conference of the international fuzzy systems association and the European society for fuzzy logic and technology (IFSA-EJSFLAT), 2015.
- Ignjatović J, Ćirić M, Stanković I (2015b) Regular fuzzy equivalences on multi-mode multi-relational fuzzy networks. In: Proceedings of the 2015 conference of the international fuzzy systems association and the European society for fuzzy logic and technology (IFSA-EUSFLAT), 2015.
- Kripke S (1959) A completeness theorem in modal logic. *J Symb Log* 24(1):1–14
- Kripke S (1963) Semantic analysis of modal logic I: normal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9:67–96
- Lerner J (2005) Role assignments. In: Brandes U, Erlebach T (eds) *Network analysis*, LNCS 3418. Springer, pp 216–252
- Liau CJ, Lin TY (2005) Reasoning about relational granulation in modal logics. In: Proceedings of the first IEEE international conference on granular computing, pp 534–558
- Lorrain F, White HC (1971) Structural equivalence of individuals in social networks. *J Math Sociol* 1:49–80
- Marti M, Metcalfe G (2014) A Hennessy-Milner property for many-valued modal logics. In: Proceedings of the tenth conference on advances in modal logic, pp 407–420
- Marx M, Masuch M (2003) Regular equivalence and dynamic logic. *Soc Netw* 25(1):51–65
- Nair PS, Sarasamma S (2007) Data mining through fuzzy social network analysis. In: Proceedings of the 26th international conference of North American fuzzy information processing society, pp 251–255
- Nardi D, Brachman RJ (2002) An introduction to description logics. In: Baader F, Calvanese D, McGuinness DL, Nardi D, Patel-Schneider PF (eds) *Description logic handbook*, Cambridge University Press, pp 5–44
- Passy S, Tinchev T (1985) Quantifiers in combinatory PDL: completeness, definability, incompleteness. In: *Fundamentals of computation theory FCT 85*, volume 199 of LNCS. Springer, pp 512–519
- Passy S, Tinchev T (1991) An essay in combinatory dynamic logic. *Inf Comput* 93:263–332
- Pawlak Z (1982) Rough sets. *Int J Inf Comput Sci* 11(15):341–356
- Pawlak Z (1991) *Rough sets—theoretical aspects of reasoning about data*. Kluwer Academic Publishers
- Pizarro N (2007) Structural identity and equivalence of individuals in social networks: beyond duality. *Int Sociol* 22(6):767–792
- Prior A (1967) *Past, present and future*. Oxford University Press
- Prior A (1968) Now. *Noûs* 2:101–119
- Scott J (2000) *Social network analysis: a handbook*, 2nd edn. SAGE publications
- Sierpinski W, Krieger C (1956) *General topology*. University of Toronto Press
- Toivonen R, Kumpula JM, Saramäki J, Onnela J-P, Kertész J, Kask K (2007) The role of edge weights in social networks: modelling structure and dynamics. In: Proceedings of SPIE 6601(1): noise and stochastics in complex systems and finance, pp B1–B8
- van der Hoek W (1992) On the semantics of graded modalities. *J Appl Non-Classical Logic* 2(1):81–123
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press
- White DR, Reitz KP (1983) Graph and semigroup homomorphisms on networks and relations. *Soc Netw* 5(1):143–234
- Wild P, Schröder L, Pattinson D, König B (2018) A van Benthem theorem for fuzzy modal logic. In: Proceedings of the 33rd annual ACM/IEEE symposium on logic in computer science (LICS), pp 909–918
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta N, Ragade R, Yager R (eds) *Advances in fuzzy set theory and applications*. North-Holland, pp 3–18
- Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst* 19:111–127

Rough Sets in Decision-Making

Roman Słowiński^{1,2}, Salvatore Greco³ and
Benedetto Matarazzo³

¹Institute of Computing Science, Poznań
University of Technology, Poznań, Poland

²Systems Research Institute, Polish Academy of
Sciences, Warsaw, Poland

³Department of Economics and Business,
University of Catania, Catania, Italy

Article Outline

Entry Outline

Glossary

Definition of the Subject and Its Importance

Introduction to Decision Support Using Rough
Set Concept

Classical Rough Set Approach to Classification
Problems of Taxonomy Type

Dominance-Based Rough Set Approach to
Ordinal Classification Problems

DRSA on a Pairwise Comparison Table for
Multiple Criteria Choice and Ranking
Problems

DRSA for Decision Under Uncertainty

Multiple Criteria Decision Analysis Using
Association Rules

Interactive Multiobjective Optimization Using
DRSA (IMO-DRSA)

An Abstract Algebra for Reasoning About
Multiple Criteria Decisions: The Bipolar
Pawlak-Brouwer-Zadeh Lattice

Conclusions

Prospective Directions of Applications

Bibliography

Entry Outline

The entry describes a knowledge discovery paradigm for multiple attribute and multiple criteria decision-making, which is based upon the concept

of rough sets. Rough set theory was introduced by Zdzisław Pawlak in 1982. Since then, it has often proved to be an excellent mathematical tool for the analysis of inconsistent information. In scientific analysis of decision problems, objects of decision (also called alternatives, actions, acts, solutions, ...) are described by attributes with preference-ordered value sets. Such attributes are called criteria. Another characteristic feature of decision problems is a monotonic relationship between evaluation of objects on criteria and their overall judgement, i.e., the better the evaluation of an object on criteria, the better its overall judgement. Objects violating this principle are called inconsistent. In order to deal with this kind of inconsistency in information describing decision problems, the authors have proposed a dominance-based rough set approach (DRSA) which generalizes the classical indiscernibility-based rough set approach (IRSA). In DRSA, the indiscernibility or tolerance relation among objects, which is used in IRSA, has been replaced by the dominance relation – the only relation uncontested in multiple criteria pairwise comparisons of objects. Given an input data about a decision problem in form of decision examples provided by a decision maker (possibly inconsistent), DRSA permits to structure this information into lower and upper approximations, corresponding to certain and possible knowledge about the decision problem. Induction algorithms run on these approximations discover, in turn, certain and possible decision rules that facilitate an understanding of the decision maker's preferences, and enable a recommendation concordant with these preferences. In this entry, DRSA is applied to the following decision problems: ordinal classification (also called multiple criteria sorting), multiple criteria choice and ranking, decision under uncertainty, and interactive multiobjective optimization. The content is structured as follows:

- Glossary
- Definition of the subject and its importance

- Introduction to decision support using rough set concept
- Classical rough set approach to classification problems of taxonomy type
- Dominance-based rough set approach to ordinal classification problems
- DRSA on a pairwise comparison table for multiple criteria choice and ranking problems
- DRSA for decision under uncertainty
- Multiple criteria decision analysis using association rules
- Interactive multiobjective optimization using DRSA (IMO-DRSA)
- Conclusions
- Future directions
- Bibliography

Glossary

Multiple attribute (or multiple criteria) decision support aims at giving the decision maker (DM) a recommendation concerning a set of *objects* A (also called alternatives, actions, acts, solutions, options, candidates, ...) evaluated from multiple points of view and called *attributes* (also called features, variables, criteria, ...).

Main categories of *multiple attribute (or multiple criteria) decision problems*:

- *Classification*, when the decision aims at assigning objects to predefined classes
- *Choice*, when the decision aims at selecting the best objects
- *Ranking*, when the decision aims at ordering objects from the best to the worst

Two kinds of *classification problems* are distinguished:

- *Taxonomy*, when the value sets of attributes and the predefined classes are not preference ordered
- *Ordinal classification* (also known as *multiple criteria sorting*), when the value sets of attributes and the predefined classes are preference ordered

- The ordinal classification is called *monotonic*, when the improvement of an object on any condition attribute, with the remaining condition attributes unchanged, cannot deteriorate the assessment of this object on the decision attribute.

Two kinds of *choice problems* are distinguished:

- *Discrete choice*, when the set of objects is finite and reasonably small to be listed
- *Multiple objective optimization*, when the set of objects is infinite and defined by constraints of a mathematical program

If value sets of attributes are preference-ordered, they are called *criteria* or *objectives*, otherwise they keep the name of attributes.

Criterion is a real-valued function g_i defined on A , reflecting a worth of objects from a particular point of view, such that in order to compare any two objects $a, b \in A$ from this point of view, it is sufficient to compare two values: $g_i(a)$ and $g_i(b)$.

Dominance: object a is non-dominated in set A (Pareto-optimal) if and only if there is no other object b in A such that b is not worse than a on all considered criteria, and strictly better on at least one criterion.

Preference model is a representation of a value system of the decision maker on the set of objects with vector evaluations.

Decision under uncertainty takes into account consequences of decision that distribute over multiple states of nature with different probability. The preference order characteristic for data describing *multiple attribute decision problems*, concerns also decision under uncertainty, where the objects correspond to acts, attributes are probabilities of a gain, and the problem consists in ordinal classification, choice, or ranking of the acts.

Rough set in universe U is an approximation of a set based on available information about objects of U . The rough approximation is composed of two ordinary sets called *lower* and *upper*

approximation. Lower approximation is a maximal subset of objects which, according to the available information, certainly belong to the approximated set, and upper approximation is a minimal subset of objects which, according to the available information, possibly belong to the approximated set. The difference between upper and lower approximation is called *boundary*.

Decision rule is a logical statement of the type “if... , then...” where the premise (condition part) specifies values assumed by one or more condition attributes, and the conclusion (decision part) specifies an overall judgement.

Definition of the Subject and Its Importance

Scientific analysis of decision problems aims at giving the decision maker (DM) a recommendation concerning a set of *objects* (also called alternatives, solutions, acts, actions, options, candidates, ...) evaluated from multiple points of view considered relevant for the problem at hand and called *attributes* (also called features, variables, criteria, ...). For example, a decision can concern:

1. Diagnosis of pathologies for a set of patients, where patients are objects of the decision, and symptoms and results of medical tests are the attributes.
2. Assignment of enterprises to classes of risk, where enterprises are objects of the decision, and financial ratio indices and other economic indicators, such as the market structure, the technology used by the enterprise, and the quality of management, are the attributes.
3. Selection of a car to be bought from among a given set of cars, where cars are objects of the decision, and maximum speed, acceleration, price, fuel consumption, comfort, color, and so on are the attributes.
4. Ordering of students applying for a scholarship, where students are objects of the decision, and scores in different disciplines are the attributes.

The following three main categories of decision problems are typically distinguished (Roy 1996):

- *Classification*, when the decision aims at assigning objects to predefined classes
- *Choice*, when the decision aims at selecting the best objects
- *Ranking*, when the decision aims at ordering objects from the best to the worst

Looking at the above examples, one can say that (1) and (2) are classification problems, (3) is a choice problem and (4) is a ranking problem.

The above categorization can be refined by distinguishing two kinds of classification problems: *taxonomy*, when the value sets of attributes and the predefined classes are not preference ordered, and *ordinal classification* (also known as *multiple criteria sorting*), when the value sets of attributes and the predefined classes are preference ordered (Greco et al. 1998a). In the above examples, (1) is a taxonomy problem and (2) is an ordinal classification problem. If value sets of attributes are preference ordered, they are called *criteria*, otherwise they keep the name of attributes. For example, in a decision regarding the selection of a car, its price is a criterion because, obviously, a low price is better than a high price. Instead, the color of a car is not a criterion but simply an attribute, because red is not intrinsically better than green. One can imagine, however, that also the color of a car could become a criterion if, for example, a DM would consider red better than green.

Introduction to Decision Support Using Rough Set Concept

Scientific support of decisions makes use of a more or less explicit model of the decision problem. The model relates the decision to the characteristics of the objects expressed by the considered attributes. Building such a model requires information about conditions and parameters of the aggregation of multi-attribute characteristics of objects. The nature of this information depends

on the adopted methodology: prices and interest rates for cost-benefit analysis, cost coefficients in objectives and technological coefficients in constraints for mathematical programming, a training set of decision examples for neural networks and machine learning, substitution rates for a value function of multi-attribute utility theory, pairwise comparisons of objects in terms of intensity of preference for the analytic hierarchy process, attribute weights, and several thresholds for ELECTRE methods, and so on (see the state-of-the-art survey Greco et al. 2016a). This information has to be provided by the DM, possibly assisted by an analyst.

Very often this information is not easily definable. For example, this is the case of the price of many immaterial goods and of the interest rates in cost-benefit analysis, or the case of the coefficients of objectives and constraints in mathematical programming models. Even if the required information is easily definable, like a training set of decision examples for neural networks, it is often processed in a way which is not clear for the DM, such that (s)he cannot see what are the exact relations between the provided information and the final recommendation. Consequently, very often the decision model is perceived by the DM as a *black box* whose result has to be accepted because the analyst's authority guarantees that the result is "right." In this context, the aspiration of the DM to find good reasons to make decision is frustrated and rises the need for a more transparent methodology in which the relation between the original information and the final recommendation is clearly shown. Such a transparent methodology searched for has been called *glass box* (Greco et al. 2008a). Its typical representative is using a training set of decision examples as the input preference information provided by the DM, and it is expressing the decision model in terms of a set of "if... then..." decision rules induced from the input information. From one side, the decision rules are explicitly related to the original information and, from the other side, they give understandable justifications for the decisions to be made.

For example, in case of a medical diagnosis problem, the decision rule approach requires as input information a set of examples of previous

diagnoses, from which some diagnostic rules are induced, such as "if there is symptom α and the test result is β , then there is pathology γ ". Each one of such rules is directly related to examples of diagnoses in the input information, where there is symptom α , test result β , and pathology γ . Moreover, the DM can verify easily that in the input information, there is no example of diagnosis where there is symptom α , test result β , but no pathology γ .

The rules induced from the input information provided in terms of exemplary decisions represent a decision model which is transparent for the DM and enables their understanding of the reasons of their past decisions. The acceptance of the rules by the DM justifies, in turn, their use for future decisions.

The induction of rules from examples is a typical approach of artificial intelligence. This explains our interest in rough set theory (Pawlak 1982, 1991) which proved to be a useful tool for analysis of vague description of decision situations (Pawlak and Słowiński 1994; Słowiński 1993). The rough set analysis aims at explaining the values of some decision attributes, playing the role of "dependent variables," by means of the values of condition attributes, playing the role of "independent variables." For example, in the above diagnostic context, data about the presence of a pathology are given by decision attributes, while data about symptoms and tests are given by condition attributes. An important advantage of the rough set approach is that it can deal with partly inconsistent examples, i.e., cases where the presence of different pathologies is associated with the presence of the same symptoms and test results. Moreover, it provides useful information about the role of particular attributes and their subsets, and prepares the ground for representation of knowledge hidden in the data by means of "if... then..." decision rules.

Classical rough set approach proposed by Pawlak (1982, 1991) cannot deal, however, with preference order in the value sets of condition and decision attributes. For this reason, the classical rough set approach can deal with only one of four decision problems listed above – classification of taxonomy type. To deal with ordinal classification,

choice, and ranking, it is necessary to generalize the classical rough set approach, so as to take into account preference orders and monotonic relationships between condition and decision attributes. This generalization, called dominance-based rough set approach (DRSA), has been proposed by Greco et al. (Greco et al. 1998a, 1999a, b, 2001a, 2002a; Słowiński et al. 2002a).

Classical Rough Set Approach to Classification Problems of Taxonomy Type

Data Table and Indiscernibility Relation

Information regarding classification examples is supplied in the form of an *information table*, whose separate rows refer to distinct *objects*, and whose columns refer to different *attributes* considered. This means that each cell of this table indicates an *evaluation* (quantitative or qualitative) of the object placed in the corresponding row by means of the attribute in the corresponding column. Formally, an information table is the 4-tuple $\mathcal{S} = \langle U, Q, V, v \rangle$, where U is a finite set of objects, called *universe*, $Q = \{q_1, \dots, q_n\}$ is a finite set of attributes, V_q is a value set of attribute q , $V = \cup_{q \in Q} V_q$, and $v: U \times Q \rightarrow V$ is a total function such that $v(x, q) \in V_q$ for each $q \in Q$, $x \in U$, called *information function*.

Therefore, each object x of U is described by a vector (string) $Des_Q(x) = [v(x, q_1), \dots, v(x, q_n)]$, called *description* of x in terms of the evaluations on the attributes from Q ; it represents the available information about x . Obviously, $x \in U$ can be described in terms of any non-empty subset $P \subseteq Q$. To every (non-empty) subset of attributes $P \subseteq Q$, there is associated an *indiscernibility relation* on U , denoted by I_P :

$$I_P = \{(x, y) \in U \times U : v(x, q) = v(y, q), \forall q \in P\}.$$

If $(x, y) \in I_P$ it is said that the objects x and y are P -indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric, and transitive). The family of all the equivalence classes of the relation I_P is

denoted by U/I_P , and the equivalence class containing object $x \in U$, by $I_P(x)$. The equivalence classes of the relation I_P are called *P-elementary sets*.

Approximations

Let $\mathcal{S}$ be an information table, X a non-empty subset of U , and $\emptyset \neq P \subseteq Q$. The *P-lower approximation* and the *P-upper approximation* of X in $\mathcal{S}$ are defined, respectively, by:

$$\begin{aligned} \underline{P}(X) &= \{x \in U : I_P(x) \subseteq X\}, \\ \overline{P}(X) &= \{x \in U : I_P(x) \cap X \neq \emptyset\}. \end{aligned}$$

The elements of $\underline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation I_P , contained in X ; the elements of $\overline{P}(X)$ are all and only those objects $x \in U$ which belong to the equivalence classes generated by the indiscernibility relation I_P , containing at least one object x belonging to X . In other words, $\underline{P}(X)$ is the largest union of the P -elementary sets included in X , while $\overline{P}(X)$ is the smallest union of the P -elementary sets containing X .

The *P-boundary* of X in $\mathcal{S}$, denoted by $Bn_P(X)$, is defined by

$$Bn_P(X) = \overline{P}(X) - \underline{P}(X).$$

The following *inclusion relation* holds: $\underline{P}(X) \subseteq X \subseteq \overline{P}(X)$.

Thus, in view of information provided by P , if an object x belongs to $\underline{P}(X)$, then it certainly belongs to set X , while if x belongs to $\overline{P}(X)$, then it possibly belongs to set X . $Bn_P(X)$ constitutes the “doubtful region” of X : nothing can be said with certainty about the membership of its elements to set X , using the subset of attributes P only.

Moreover, the following *complementarity relation* is satisfied:

$$\underline{P}(X) = U - \overline{P}(U - X).$$

If the P -boundary of set X is empty, $Bn_P(X) = \emptyset$, then X is an ordinary (exact) set with respect to P , that is, it may be expressed as

union of a certain number of P -elementary sets; otherwise, if $Bn_P(X) \neq \emptyset$, set X is an approximate (rough) set with respect to P and may be characterized by means of the approximations $\underline{P}(X)$ and $\overline{P}(X)$. The family of all sets $X \subseteq U$ having the same P -lower and P -upper approximations is called the *rough set*.

The *quality* of the approximation of set X by means of the attributes from P is defined as

$$\gamma_P(X) = \frac{|\underline{P}(X)|}{|X|},$$

such that $0 \leq \gamma_P(X) \leq 1$. The quality $\gamma_P(X)$ represents the relative frequency of the objects correctly classified using the attributes from P .

The definition of approximations of a set $X \subseteq U$ can be extended to a classification, i.e., a partition $Y = \{Y_1, \dots, Y_m\}$ of U . The P -lower and P -upper approximations of Y in $\mathcal{S}$ are defined by sets $\underline{P}Y = \{\underline{P}Y_1, \dots, \underline{P}Y_m\}$ and $\overline{P}Y = \{\overline{P}Y_1, \dots, \overline{P}Y_m\}$, respectively. The coefficient

$$\gamma_P(Y) = \frac{\sum_{i=1}^m |\underline{P}Y_i|}{|U|}$$

is called *quality of the approximation of classification* Y by set of attributes P , or in short, *quality of classification*. It expresses the ratio of all P -correctly classified objects to all objects in the system.

Dependence and Reduction of Attributes

An issue of great practical importance is the reduction of “superfluous” attributes in an information table. Superfluous attributes can be eliminated without deteriorating the information contained in the original table.

Let $P \subseteq Q$ and $p \in P$. It is said that attribute p is *superfluous* in P with respect to classification Y if $\underline{P}(Y) = \underline{(P - \{p\})}(Y)$; otherwise, p is *indispensable* in P . The subset of Q containing all the indispensable attributes is known as the *core*.

Given classification Y , any minimal (with respect to inclusion) subset $P \subseteq Q$, such that $\underline{P}(Y) = \underline{Q}(Y)$, is called *reduct*. It specifies a minimal subset P of

Q which keeps the quality of classification at the same level as the whole set of attributes, i.e., $\gamma_P(Y) = \gamma_Q(Y)$. In other words, the attributes that do not belong to the reduct are superfluous with respect to the classification Y of objects from U .

More than one reduct may exist in an information table and their intersection gives the *core*.

Decision Table and Decision Rules

In the information table describing examples of classification, the attributes of set Q are divided into *condition* attributes (set $C \neq \emptyset$) and *decision* attributes (set $D \neq \emptyset$), $C \cup D = Q$ and $C \cap D = \emptyset$. Such an information table is called a *decision table*. The decision attributes induce a partition of U deduced from the indiscernibility relation I_D in a way that is independent of the condition attributes. D -elementary sets are called *decision classes*, denoted by Cl_t , $t = 1, \dots, m$. The partition of U into decision classes is called *classification* $\mathbf{CI} = \{Cl_1, \dots, Cl_m\}$. There is a tendency to reduce the set C while keeping all important relationships between C and D , in order to make decisions on the basis of a smaller amount of information. When the set of condition attributes is replaced by one of its reducts, the quality of approximation of the classification induced by the decision attributes is not deteriorating.

Since the aim is to underline the functional dependencies between condition and decision attributes, a decision table may also be seen as a set of *decision rules*. These are logical statements of the type “if... then...” where the premise (condition part) specifies values assumed by one or more condition attributes (description of C -elementary sets) and the conclusion (decision part) specifies an assignment to one or more decision classes. Therefore, the syntax of a rule is the following:

“if $v(x, q_1) = r_{q_1}$ and $v(x, q_2) = r_{q_2}$ and ...
 $v(x, q_p) = r_{q_p}$, then x belongs to decision class
 Cl_{j_1} or Cl_{j_2} or ... Cl_{j_k} ,”

where $\{q_1, q_2, \dots, q_p\} \subseteq C$, $(r_{q_1}, r_{q_2}, \dots, r_{q_p}) \in V_{q_1} \times V_{q_2} \times \dots \times V_{q_p}$ and $Cl_{j_1}, Cl_{j_2}, \dots, Cl_{j_k}$ are some decision classes of the considered

classification *CI*. If the consequence is univocal, i.e., $k = 1$, then the rule is *univocal*, otherwise it is *approximate*.

An object $x \in U$ *supports* decision rule r if its description is matching both the condition part and the decision part of the rule. The decision rule r *covers* object x if it matches the condition part of the rule. Each decision rule is characterized by its *strength*, defined as the number of objects supporting the rule. In the case of approximate rules, the strength is calculated for each possible decision class separately.

If a univocal rule is supported by objects from the lower approximation of the corresponding decision class only, then the rule is called *certain* or *deterministic*. If, however, a univocal rule is supported by objects from the upper approximation of the corresponding decision class only, then the rule is called *possible* or *probabilistic*. Approximate rules are supported, in turn, only by objects from the boundaries of the corresponding decision classes.

Procedures for generation of decision rules from a decision table use an inductive learning principle. The objects are considered as examples of classification. In order to induce a decision rule with a univocal and certain conclusion about assignment of an object to decision class X , the examples belonging to the C -lower approximation of X are called *positive* and all the others *negative*. Analogously, in case of a possible rule, the examples belonging to the C -upper approximation of X are positive and all the others negative. Possible rules are characterized by a coefficient, called *confidence*, telling to what extent the rule is consistent, i.e., what is the ratio of the number of positive examples supporting the rule to the number of examples belonging to set X according to decision attributes. Finally, in case of an approximate rule, the examples belonging to the C -boundary of X are positive and all the others negative. A decision rule is called *minimal* if removing any attribute from the condition part gives a rule covering also negative objects.

The existing induction algorithms use one of the following strategies (Stefanowski 1998):

- (a) Generation of a *minimal* representation, i.e., minimal set of rules covering all objects from a decision table
- (b) Generation of an *exhaustive* representation, i.e., all rules for a given decision table
- (c) Generation of a *characteristic* representation, i.e., a set of rules covering relatively many objects, however, not necessarily all objects from a decision table.

Explanation of the Classical Rough Set Approach by an Example

Suppose that one wants to describe the classification of basic traffic signs to a novice. There are three main classes of traffic signs corresponding to:

- Warning (W)
- Interdiction (I)
- Order (O)

Then, these classes may be distinguished by such attributes as the shape (S) and the principal color (PC) of the sign. Finally, one can consider a few examples of traffic signs, like those shown in Table 1. These are:

- (a) Sharp right turn.
- (b) Speed limit of 50 km/h.
- (c) No parking.
- (d) Go ahead.

Rough Sets in Decision-Making, Table 1 Examples of traffic signs described by S and PC

Traffic sign	Shape (S)	Primary Color (PC)	Class
a) 	triangle	yellow	W
b) 	circle	white	I
c) 	circle	blue	I
d) 	circle	blue	O

The rough set approach is used here to build a model of classification of traffic signs to classes W, I, O on the basis of attributes S and PC. This is a typical problem of taxonomy.

One can remark that the sets of signs indiscernible by “Class” are:

$$W = \{a\}_{Class}, \quad I = \{b, c\}_{Class}, \quad O = \{d\}_{Class},$$

and the sets of signs indiscernible by S and PC are as follows:

$$\{a\}_{S,PC}, \quad \{b\}_{S,PC}, \quad \{c, d\}_{S,PC}.$$

The above elementary sets are generated, on the one hand, by decision attribute “Class” and, on the other hand, by condition attributes S and PC. The elementary sets of signs indiscernible by “Class” are denoted by $\{\cdot\}_{Class}$ and those by S and PC are denoted by $\{\cdot\}_{S,PC}$. Notice that $W = \{a\}_{Class}$ is characterized precisely by $\{a\}_{S,PC}$. In order to characterize $I = \{b, c\}_{Class}$ and $O = \{d\}_{Class}$, one needs $\{b\}_{S,PC}$ and $\{c, d\}_{S,PC}$; however, only $\{b\}_{S,PC}$ is included in $I = \{b, c\}_{Class}$ while $\{c, d\}_{S,PC}$ has a non-empty intersection with both $I = \{b, c\}_{Class}$ and $O = \{d\}_{Class}$. It follows, from this characterization, that by using condition attributes S and PC, one can characterize class W precisely, while classes I and O can only be characterized approximately:

- Class W includes sign a certainly and possibly no other sign than a .
- Class I includes sign b certainly and possibly signs b, c , and d .
- Class O includes no sign certainly and possibly signs c and d .

The terms “*certainly*” and “*possibly*” refer to the absence or presence of ambiguity between the description of signs by S and PC from the one side, and by “Class,” from the other side. In other words, using description of signs by S and PC, one can say that all signs from elementary sets $\{\cdot\}_{S,PC}$ included in elementary sets $\{\cdot\}_{Class}$ belong certainly to the corresponding class,

while all signs from elementary sets $\{\cdot\}_{S,PC}$ having a non-empty intersection with elementary sets $\{\cdot\}_{Class}$ belong to the corresponding class only possibly. The two sets of certain and possible signs are, respectively, the *lower* and *upper approximation* of the corresponding class by attributes S and PC:

$$\begin{aligned} \text{lower_approx}_{S,PC}(W) &= \{a\}, \\ \text{upper_approx}_{S,PC}(W) &= \{a\}, \\ \text{lower_approx}_{S,PC}(I) &= \{b\}, \\ \text{upper_approx}_{S,PC}(I) &= \{b, c, d\}, \\ \text{lower_approx}_{S,PC}(O) &= \emptyset, \\ \text{upper_approx}_{S,PC}(O) &= \{c, d\}. \end{aligned}$$

The *quality of approximation* of the classification by attributes S and PC is equal to the number of all the signs in the lower approximations divided by the number of all the signs in the table, i.e., $\frac{1}{2}$.

One way to increase the quality of the approximation is to add a new attribute so as to decrease the ambiguity. Let us introduce the secondary color (SC) as a new condition attribute. The new situation is now shown in Table 2.

As one can see, the sets of signs indiscernible by S, PC, and SC, i.e., the elementary sets $\{\cdot\}_{S,PC,SC}$, are now:

$$\{a\}_{S,PC,SC}, \quad \{b\}_{S,PC,SC}, \quad \{c\}_{S,PC,SC}, \quad \{d\}_{S,PC,SC}.$$

Rough Sets in Decision-Making, Table 2 Examples of traffic signs described by S, PC, and SC

Traffic sign	Shape (S)	Primary Color (PC)	Secondary color (SC)	Class
a) 	triangle	yellow	red	W
b) 	circle	white	red	I
c) 	circle	blue	red	I
d) 	circle	blue	white	O

It is worth noting that the elementary sets are finer than before and this enables the ambiguity to be eliminated. Consequently, the quality of approximation of the classification by attributes S, PC, and SC is now equal to 1.

A natural question occurring here is to ask if, indeed, all three attributes are necessary to characterize precisely the classes W, I, and O. When attribute S or attribute PC is eliminated from the description of the signs, the elementary sets $\{\cdot\}_{PC,SC}$ or $\{\cdot\}_{S,SC}$ are defined, respectively, as follows:

$$\{a\}_{PC,SC}, \{b\}_{PC,SC}, \{c\}_{PC,SC}, \{d\}_{PC,SC}, \\ \{a\}_{S,SC}, \{b,c\}_{S,SC}, \{d\}_{S,SC}.$$

Using any one of the above elementary sets, it is possible to characterize (approximate) classes W, I, and O with the same quality (equal to 1) as it is when using the elementary sets $\{\cdot\}_{S,PC,SC}$ (i.e., those generated by the complete set of three condition attributes). Thus, the answer to the above question is that the three condition attributes are not necessary to characterize precisely the classes W, I, and O. It is, in fact, sufficient to use either PC and SC or S and SC. The subsets of condition attributes $\{PC, SC\}$ and $\{S, SC\}$ are called *reducts* of $\{S, PC, SC\}$ because they have this property. Note that the identification of reducts enables us to reduce attributes about the signs from the table to those which are relevant.

Other useful information can be generated from the identification of reducts by taking their intersection. This is called the *core*. In our example, the core contains attribute SC. This tells us that it is clearly an *indispensable* attribute, i.e., it cannot be eliminated from the description of the signs without decreasing the quality of the approximation. Note that other attributes from the reducts (i.e., S and PC) are *exchangeable*. If there happened to be some other attributes which were not included in any reduct, then they would be *superfluous*, i.e., they would not be useful at all in the characterization of the classes W, I, and O.

If, however, column S or PC is eliminated from Table 2, then the resulting table is not a minimal representation of knowledge about the classification

of the four traffic signs. Note that, in order to characterize class W in Table 2, it is sufficient to use the condition “S = triangle.” Moreover, class I is characterized by two conditions (“S = circle” and “SC = red”) and class O is characterized by the condition “SC = white.” Thus, the minimal representation of this information system requires only four conditions (rather than the eight conditions that are presented in Table 2 with either column S or PC eliminated). This representation corresponds to the following set of *decision rules* which may be seen as classification model discovered in the data set contained in Table 2 (in the braces, there are symbols of signs covered by the corresponding rule):

rule #1: if S = triangle, then Class = W {a}
 rule #2: if S = circle
 and SC = red, then Class = I {b, c}
 rule #3: if SC = white, then Class = O {d}

This is not the only representation, because an alternative set of rules is:

rule #1': if PC = yellow, then Class = W {a}
 rule #2': if PC = white, then Class = I {b}
 rule #3': if PC = white
 and SC = red, then Class = I {c}
 rule #4': if SC = white, then Class = O {d}

It is interesting to come back to Table 1 and to ask what decision rules represent this information system. As the description of the four signs by S and PC is not sufficient to characterize exactly all the classes, it is not surprising that not all the rules will have a nonambiguous decision. Indeed, the following decision rules can be induced:

rule #1'': if S = triangle, then Class = W {a}
 rule #2'': if PC = white, then Class = I {b}
 rule #3'': if PC = blue, then Class = I or O {c, d}

Note that these rules can be induced from the lower approximations of classes W and I, and

from the set called the *boundary* of both I and O. Indeed, for exact rule #1'', the supporting example is in $\text{lower_appx}_{S,PC}(W) = \{a\}$; for exact rule #2'', it is in $\text{lower_appx}_{S,PC}(I) = \{b\}$; and the supporting examples for approximate rule #3'' are in the boundary of classes I and O, defined as:

$$\begin{aligned} \text{boundary}_{S,PC}(I) &= \text{upper_appx}_{S,PC}(I) \\ &\quad - \text{lower_appx}_{S,PC}(I) \\ &= \{c, d\}, \end{aligned}$$

$$\begin{aligned} \text{boundary}_{S,PC}(O) &= \text{upper_appx}_{S,PC}(O) \\ &\quad - \text{lower_appx}_{S,PC}(O) \\ &= \{c, d\}. \end{aligned}$$

As a result of the approximate characterization of classes W, I, and O by S and PC, an approximate representation in terms of decision rules is obtained. Since the quality of the approximation is $\frac{1}{2}$, exact rules (#1'' and #2'') cover one half of the examples and the other half is covered by the approximate rule (#3''). Now, the quality of approximation by S and SC, or by PC and SC, was equal to 1, so all examples were covered by exact rules (#1 to #3, or #1' to #4', respectively).

One can see, from this simple example, that the rough set analysis of data included in an information system provides some useful information. In particular, the following results are obtained:

- A characterization of decision classes in terms of chosen condition attributes through lower and upper approximation.
- A measure of the quality of approximation which indicates how good the chosen set of attributes is for approximation of the classification.
- The reducts of condition attributes including all relevant attributes. At the same time, superfluous and exchangeable attributes are also identified.
- The core composed of indispensable attributes.
- A set of decision rules which is induced from the lower and upper approximations of the

decision classes. This constitutes a classification model for a given information system.

Dominance-Based Rough Set Approach to Ordinal Classification Problems

Dominance-Based Rough Set Approach (DRSA)

Dominance-based rough set approach (DRSA) has been proposed by the authors to handle background knowledge about ordinal evaluations of objects from a universe, and about monotonic relationships between these evaluations, e.g., "the larger the mass and the smaller the distance, the larger the gravity" or "the greater the debt of a firm, the greater its risk of failure". Such a knowledge is typical for data describing various phenomena. It is also characteristic for data concerning multiple criteria decision or decision under uncertainty, where the order of value sets of condition and decision attributes corresponds to increasing or decreasing preference. In case of multiple criteria decision, the condition attributes are called *criteria*.

Let us consider a decision table including a finite universe of objects (solutions, alternatives, actions) U evaluated on a finite set of criteria $F = \{f_1, \dots, f_n\}$, and on a single decision attribute d . The set of the indices of criteria is denoted by $I = \{1, \dots, n\}$. Without loss of generality, $f_i : U \rightarrow \mathfrak{R}$ for each $i = 1, \dots, n$, and, for all objects $x, y \in U$, $f_i(x) \geq f_i(y)$ means that "x is at least as good as y with respect to criterion i ," which is denoted by $x \succeq_i y$. Therefore, it is supposed that $\succeq_i$ is a complete preorder, i.e., a strongly complete and transitive binary relation, defined on U on the basis of evaluations $f_i(\cdot)$. Furthermore, decision attribute d makes a partition of U into a finite number of decision classes, $CI = \{Cl_1, \dots, Cl_m\}$, such that each $x \in U$ belongs to one and only one class Cl_t , $t = 1, \dots, m$. It is assumed that the classes are preference ordered, i.e., for all $r, s = 1, \dots, m$, such that $r > s$, the objects from Cl_r are preferred to the objects from Cl_s . More formally, if $\succeq$ is a *comprehensive weak preference relation* on U , i.e., if for all $x, y \in U$,

$x \succeq y$ reads “ x is at least as good as y ,” then it is supposed that

$$[x \in Cl_r, y \in Cl_s, r > s] \Rightarrow x \succ y,$$

where $x \succ y$ means $x \succeq y$ and *not* $y \succeq x$.

The above assumptions are typical for consideration of an ordinal classification (or multiple criteria sorting) problem. Indeed, the decision table characterized above includes examples of ordinal classification which constitute an input *preference information* to be analyzed using DRSA.

The sets to be approximated are called *upward union* and *downward union* of decision classes, respectively:

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s, \quad Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s, \quad t = 1, \dots, m.$$

The statement $x \in Cl_t^{\geq}$ reads “ x belongs to at least class Cl_t ,” while $x \in Cl_t^{\leq}$ reads “ x belongs to at most class Cl_t .” Let us remark that $Cl_1^{\geq} = Cl_m^{\leq} = U$, $Cl_m^{\geq} = Cl_m^{\leq}$ and $Cl_1^{\leq} = Cl_1^{\geq}$. Furthermore, for $t = 2, \dots, m$,

$$Cl_{t-1}^{\leq} = U - Cl_t^{\geq} \quad \text{and} \quad Cl_t^{\geq} = U - Cl_{t-1}^{\leq}.$$

The key idea of DRSA is representation (approximation) of upward and downward unions of decision classes, by *granules of knowledge* generated by criteria. These granules are *dominance cones* in the criteria values space.

x dominates y with respect to set of criteria $P \subseteq I$ (shortly, x *P-dominates* y), denoted by $x D_P y$, if for every criterion $i \in P$, $f_i(x) \geq f_i(y)$. The relation of *P-dominance* is reflexive and transitive, i.e., it is a partial preorder.

Given a set of criteria $P \subseteq I$ and $x \in U$, the granules of knowledge used for approximation in DRSA are:

- A set of objects dominating x , called *P-dominating set*,

$$D_P^+(x) = \{y \in U : y D_P x\}.$$

- A set of objects dominated by x , called *P-dominated set*,

$$D_P^-(x) = \{y \in U : x D_P y\}.$$

Let us recall that the dominance principle requires that an object x dominating object y on all considered criteria (i.e., x having evaluations at least as good as y on all considered criteria) should also dominate y on the decision (i.e., x should be assigned to at least as good decision class as y). Objects satisfying the dominance principle are called *consistent*, and those which are violating the dominance principle are called *inconsistent*.

The *P-lower approximation* of $Cl_t^{\geq}$, denoted by $\underline{P}(Cl_t^{\geq})$, and the *P-upper approximation* of $Cl_t^{\geq}$, denoted by $\overline{P}(Cl_t^{\geq})$, are defined as follows ($t = 2, \dots, m$):

$$\begin{aligned} \underline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\}, \\ \overline{P}(Cl_t^{\geq}) &= \{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\}. \end{aligned}$$

Analogously, one can define the *P-lower approximation* and the *P-upper approximation* of $Cl_t^{\leq}$ as follows ($t = 1, \dots, m-1$):

$$\begin{aligned} \underline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\}, \\ \overline{P}(Cl_t^{\leq}) &= \{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\}. \end{aligned}$$

The *P-lower* and *P-upper* approximations so defined satisfy the following *inclusion properties*, for all $P \subseteq I$:

$$\begin{aligned} \underline{P}(Cl_t^{\geq}) &\subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq}), \quad t = 2, \dots, m, \\ \underline{P}(Cl_t^{\leq}) &\subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq}), \quad t = 1, \dots, m-1. \end{aligned}$$

The *P-lower* and *P-upper* approximations of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ have an important *complementarity property*, according to which,

$$\begin{aligned} \underline{P}(Cl_t^{\geq}) &= U - \overline{P}(Cl_{t-1}^{\leq}) \quad \text{and} \\ \overline{P}(Cl_t^{\geq}) &= U - \underline{P}(Cl_{t-1}^{\leq}), \quad t = 2, \dots, m, \\ \underline{P}(Cl_t^{\leq}) &= U - \overline{P}(Cl_{t+1}^{\geq}) \quad \text{and} \\ \overline{P}(Cl_t^{\leq}) &= U - \underline{P}(Cl_{t+1}^{\geq}), \quad t = 1, \dots, m-1. \end{aligned}$$

The P -boundary of $Cl_t^{\geq}$ and $Cl_t^{\leq}$, denoted by $Bn_P(Cl_t^{\geq})$ and $Bn_P(Cl_t^{\leq})$, respectively, are defined as follows:

$$\begin{aligned} Bn_P(Cl_t^{\geq}) &= \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq}), \quad t = 2, \dots, m, \\ Bn_P(Cl_t^{\leq}) &= \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq}), \quad t = 1, \dots, m-1. \end{aligned}$$

Due to the above complementarity property, $Bn_P(Cl_t^{\geq}) = Bn_P(Cl_{t-1}^{\leq})$, for $t = 2, \dots, m$.

For every $P \subseteq C$, the *quality of approximation* of the ordinal classification $\mathbf{CI}$ by a set of criteria P is defined as the ratio of the number of objects P -consistent with the dominance principle and the number of all the objects in U . Since the P -consistent objects are those which do not belong to any P -boundary $Bn_P(Cl_t^{\geq})$, $t = 2, \dots, m$, or $Bn_P(Cl_t^{\leq})$, $t = 1, \dots, m-1$, the quality of approximation of the ordinal classification $\mathbf{CI}$ by a set of criteria P can be written as

$$\begin{aligned} \gamma_P(\mathbf{CI}) &= \frac{\left| U - \left(\bigcup_{t=2, \dots, m} Bn_P(Cl_t^{\geq}) \right) \right|}{|U|} \\ &= \frac{\left| U - \left(\bigcup_{t=1, \dots, m-1} Bn_P(Cl_t^{\leq}) \right) \right|}{|U|}. \end{aligned}$$

$\gamma_P(\mathbf{CI})$ can be seen as a degree of consistency of the objects from U , where P is the set of criteria and $\mathbf{CI}$ is the considered ordinal classification.

Each minimal (with respect to inclusion) subset $P \subseteq C$ such that $\gamma_P(\mathbf{CI}) = \gamma_C(\mathbf{CI})$ is called a *reduct* of $\mathbf{CI}$ and is denoted by $RED_{\mathbf{CI}}$. Let us remark that for a given set U , one can have more than one reduct. A fairly efficient algorithm for generation of reducts in DRSA has been proposed in Susmaga et al. (2000). The intersection of all reducts is called the *core* and is denoted by $CORE_{\mathbf{CI}}$. Criteria in $CORE_{\mathbf{CI}}$ cannot be removed from consideration without deteriorating the quality of approximation. This means that, in set C , there are three categories of criteria:

- *Indispensable* criteria included in the core
- *Exchangeable* criteria included in some reducts, but not in the core
- *Redundant* criteria, neither indispensable nor exchangeable, and thus not included in any reduct

The dominance-based rough approximations of upward and downward unions of decision classes can serve to induce a generalized description of objects in terms of “if ..., then ...” decision rules. For a given upward or downward union of classes, $Cl_t^{\geq}$ or $Cl_s^{\leq}$, the decision rules induced under a hypothesis that objects belonging to $\underline{P}(Cl_t^{\geq})$ or $\underline{P}(Cl_s^{\leq})$ are positive examples, and all the others are negative, suggest a *certain* assignment to “class Cl_t or better,” or to “class Cl_s or worse,” respectively. On the other hand, the decision rules induced under a hypothesis that objects belonging to $\overline{P}(Cl_t^{\geq})$ or $\overline{P}(Cl_s^{\leq})$ are positive examples, and all the others are negative, suggest a *possible* assignment to “class Cl_t or better,” or to “class Cl_s or worse,” respectively. Finally, the decision rules induced under a hypothesis that objects belonging to the intersection $\overline{P}(Cl_s^{\leq}) \cap \overline{P}(Cl_t^{\geq})$ are positive examples, and all the others are negative, suggest an *approximate* assignment to some classes between Cl_s and Cl_t ($s < t$).

In the case of preference ordered description of objects, set U is composed of examples of ordinal classification. Then, it is meaningful to consider the following five types of decision rules:

1. *Certain $D_{\geq}$ -decision rules*, providing lower profile descriptions for objects belonging to $\underline{P}(Cl_t^{\geq})$:
 if $f_{i_1}(x) \geq r_{i_1}$ and ... and $f_{i_p}(x) \geq r_{i_p}$, then
 $x \in Cl_t^{\geq}$, $\{i_1, \dots, i_p\} \subseteq I$, $t = 2, \dots, m$,
 $r_{i_1}, \dots, r_{i_p} \in \mathfrak{R}$.
2. *Possible $D_{\geq}$ -decision rules*, providing lower profile descriptions for objects belonging to $\overline{P}(Cl_t^{\geq})$:
 if $f_{i_1}(x) \geq r_{i_1}$ and ... and $f_{i_p}(x) \geq r_{i_p}$, then
 x possibly belongs to $Cl_t^{\geq}$, $\{i_1, \dots, i_p\} \subseteq I$,
 $t = 2, \dots, m$, $r_{i_1}, \dots, r_{i_p} \in \mathfrak{R}$.
3. *Certain $D_{\leq}$ -decision rules*, providing upper profile descriptions for objects belonging to $\underline{P}(Cl_t^{\leq})$:
 if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then
 $x \in Cl_t^{\leq}$, $\{i_1, \dots, i_p\} \subseteq I$, $t = 1, \dots, m-1$,
 $r_{i_1}, \dots, r_{i_p} \in \mathfrak{R}$.

4. Possible $D_{\leq}$ -decision rules, providing upper profile descriptions for objects belonging to $\bar{P}(Cl_t^{\leq})$:

if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then x possibly belongs to $Cl_t^{\leq}$, $\{i_1, \dots, i_p\} \subseteq I$, $t = 1, \dots, m-1$, $r_{i_1}, \dots, r_{i_p} \in \mathfrak{R}$.

5. Approximate $D_{\geq\leq}$ -decision rules, providing simultaneously lower and upper profile descriptions for objects belonging to $Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$, without possibility of discerning to which class:

if $f_{i_1}(x) \geq r_{i_1}$ and ... and $f_{i_k}(x) \geq r_{i_k}$ $f_{i_{k+1}}(x) \leq r_{i_{k+1}}$... and $f_{i_p}(x) \leq r_{i_p}$, then $x \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$, $\{i_1, \dots, i_p\} \subseteq I$, $s, t \in \{1, \dots, m\}$, $s < t$, $r_{i_1}, \dots, r_{i_p} \in \mathfrak{R}$.

In the premise of a $D_{\geq\leq}$ -decision rule, there can be " $f_i(x) \geq r_i$ " and " $f_i(x) \leq r'_i$ " where $r_i \leq r'_i$, for the same $i \in I$. Moreover, if $r_i = r'_i$, the two conditions boil down to " $f_i(x) = r_i$."

Since a decision rule is a kind of implication, a *minimal* rule is understood as an implication such that there is no other implication with the premise of at least the same weakness (in other words, a rule using a subset of elementary conditions or/and weaker elementary conditions) and the conclusion of at least the same strength (in other words, a $D_{\geq}$ - or a $D_{\leq}$ -decision rule assigning objects to the same union or sub-union of classes, or a $D_{\geq\leq}$ -decision rule assigning objects to the same or larger set of classes).

The rules of type (1) and (3) represent certain knowledge extracted from data (examples of ordinal classification), while the rules of type (2) and (4) represent possible knowledge; the rules of type (5) represent doubtful knowledge, because they are supported by inconsistent objects only.

Moreover, the rules of type (1) and (3) are *exact* if they do not cover negative examples, and they are *probabilistic* otherwise. In the latter case, each rule is characterized by a *confidence ratio*, representing the probability that an object matching the premise of the rule also matches its conclusion.

Given a certain or possible $D_{\geq}$ -decision rule $r \equiv$ "if $f_{i_1}(x) \geq r_{i_1}$ and ... and $f_{i_p}(x) \geq r_{i_p}$, then $x \in Cl_t^{\geq}$," an object $y \in U$ supports r if $f_{i_1}(y) \geq r_{i_1}$ and ... and $f_{i_p}(y) \geq r_{i_p}$. Moreover, object $y \in U$ supporting decision rule r is a *base* of this rule

$f_{i_1}(y) = r_{i_1}$ and ... and $f_{i_p}(y) = r_{i_p}$. Similar definitions hold for certain or possible $D_{\leq}$ -decision rules and approximate $D_{\geq\leq}$ -decision rules. A decision rule having at least one base is called *robust*. Identification of supporting objects and bases of robust rules is important for interpretation of the rules in multiple criteria decision analysis perspective. The ratio of the number of objects supporting a rule and the number of all considered objects is called *relative support* of a rule. The relative support and the confidence ratio are basic characteristics of a rule; however, some *Bayesian confirmation measures* reflect much better the attractiveness of a rule (Greco et al. 2004c, 2016c).

A set of decision rules is *complete* if it covers all considered objects (examples of ordinal classification) in such a way that consistent objects are reassigned to their original classes, and inconsistent objects are assigned to clusters of classes referring to this inconsistency. We call *minimal* each set of decision rules that is complete and nonredundant, i.e., exclusion of any rule from this set makes it incomplete.

Note that the syntax of decision rules induced from rough approximations defined using dominance cones is using consistently this type of granules. Each condition profile defines a dominance cone in n -dimensional condition space $\mathfrak{R}^n$, and each decision profile defines a dominance cone in one-dimensional decision space $\{1, \dots, m\}$. In both cases, the cones are positive for $D_{\geq}$ -rules and negative for $D_{\leq}$ -rules.

Let also remark that dominance cones corresponding to condition profiles can originate in any point of $\mathfrak{R}^n$, without the risk of their being too specific. Thus, contrary to traditional granular computing, the condition space $\mathfrak{R}^n$ need not be discretized.

Procedures for rule induction from dominance-based rough approximations have been proposed in Greco et al. (2001c), Susmaga et al. (2000) and improved in Błaszczyński et al. (2011).

In Giove et al. (2002), a new methodology for the induction of monotonic decision trees from dominance-based rough approximations of preference ordered decision classes has been proposed.

Application of DRSA to decision-related problems goes far beyond ordinal classification. In Greco et al. (2006a), DRSA has been used for decision support involving multiple decision makers, and in Greco et al. (2006b), DRSA has been applied to case-based reasoning. The following sections present applications of DRSA to multiple criteria choice and ranking, to decision under uncertainty and to interactive multiobjective optimization. The surveys (Greco et al. 2004b, 2007b, 2016b; Słowiński et al. 2007, 2014) include other applications of DRSA.

Example Illustrating DRSA in the Context of Ordinal Classification

This subsection presents a didactic example which illustrates the main concepts of DRSA. Let us consider the following ordinal classification problem. Students of a college must obtain an overall evaluation on the basis of their achievements in Mathematics, Physics, and Literature. The three subjects are clearly criteria (condition attributes) and the comprehensive evaluation is a decision attribute. For simplicity, the value sets of the criteria and of the decision attribute are the same, and they are composed of three values: bad, medium, and good. The preference order of these values is obvious. Thus, there are three preference ordered decision classes, so the problem belongs to the category of ordinal classification. In order to build a preference model of the jury, DRSA is used to analyze a set of exemplary evaluations of students provided by the jury. These examples of ordinal classification constitute an input preference information presented as decision table in Table 3.

Note that the dominance principle obviously applies to the examples of ordinal classification, since an improvement of a student's score on one of three criteria, with other scores unchanged, should not worsen the student's overall evaluation, but rather improve it.

Observe that student S_1 has not worse evaluations than student S_2 on all the considered criteria; however, the overall evaluation of S_1 is worse than the overall evaluation of S_2 . This contradicts the dominance principle, so the two examples of ordinal classification, and only those, are inconsistent. One can expect that the quality of approximation of the ordinal classification represented by examples in Table 3 will be equal to 0.75.

One can observe that in result of reducing the set of considered criteria, i.e., the set of considered subjects, some new inconsistencies can occur. For example, removing from Table 3 the evaluation on Literature, one obtains Table 4, where S_1 is inconsistent not only with S_2 but also with S_3 and S_5 . In fact, student S_1 has not worse evaluations than students S_2 , S_3 , and S_5 on all the considered criteria (Mathematics and Physics); however, the overall evaluation of S_1 is worse than the overall evaluation of S_2 , S_3 , and S_5 .

Observe, moreover, that removing from Table 3 the evaluations on Mathematics, one obtains Table 5, where no new inconsistencies occur, comparing to Table 3.

Similarly, after removing from Table 3 the evaluations on Physics, one obtains Table 6, where no new inconsistencies occur, comparing to Table 3.

The fact that no new inconsistency occurs when Mathematics or Physics is removed, means that the

Rough Sets in Decision-Making, Table 3 Exemplary evaluations of students (examples of ordinal classification)

Student	Mathematics	Physics	Literature	Overall evaluation
S_1	Good	Medium	Bad	Bad
S_2	Medium	Medium	Bad	Medium
S_3	Medium	Medium	Medium	Medium
S_4	Good	Good	Medium	Good
S_5	Good	Medium	Good	Good
S_6	Good	Good	Good	Good
S_7	Bad	Bad	Bad	Bad
S_8	Bad	Bad	Medium	Bad

Rough Sets in Decision-Making, Table 4 Exemplary evaluations of students excluding Literature

Student	Mathematics	Physics	Overall evaluation
<i>S</i> 1	Good	Medium	Bad
<i>S</i> 2	Medium	Medium	Medium
<i>S</i> 3	Medium	Medium	Medium
<i>S</i> 4	Good	Good	Good
<i>S</i> 5	Good	Medium	Good
<i>S</i> 6	Good	Good	Good
<i>S</i> 7	Bad	Bad	Bad
<i>S</i> 8	Bad	Bad	Bad

Rough Sets in Decision-Making, Table 5 Exemplary evaluations of students excluding Mathematics

Student	Physics	Literature	Overall evaluation
<i>S</i> 1	Medium	Bad	Bad
<i>S</i> 2	Medium	Bad	Medium
<i>S</i> 3	Medium	Medium	Medium
<i>S</i> 4	Good	Medium	Good
<i>S</i> 5	Medium	Good	Good
<i>S</i> 6	Good	Good	Good
<i>S</i> 7	Bad	Bad	Bad
<i>S</i> 8	Bad	Medium	Bad

Rough Sets in Decision-Making, Table 6 Exemplary evaluations of students excluding Physics

Student	Mathematics	Literature	Overall evaluation
<i>S</i> 1	Good	Bad	Bad
<i>S</i> 2	Medium	Bad	Medium
<i>S</i> 3	Medium	Medium	Medium
<i>S</i> 4	Good	Medium	Good
<i>S</i> 5	Good	Good	Good
<i>S</i> 6	Good	Good	Good
<i>S</i> 7	Bad	Bad	Bad
<i>S</i> 8	Bad	Medium	Bad

subsets of criteria {Physics, Literature} or {Mathematics, Literature} contain sufficient information to represent the overall evaluation of students with the same quality of approximation as using the complete set of three criteria. This is not the case,

however, for the subset {Mathematics, Physics}. Observe, moreover, that subsets {Physics, Literature} and {Mathematics, Literature} are minimal, because no other criterion can be removed without new inconsistencies occur. Thus, {Physics, Literature} and {Mathematics, Literature} are the reducts of the complete set of criteria {Mathematics, Physics, Literature}. Since Literature is the only criterion which cannot be removed from any reduct without introducing new inconsistencies, it constitutes the core, i.e., the set of indispensable criteria. The core is, of course, the intersection of all reducts, i.e., in our example:

$$\{\text{Literature}\} = \{\text{Physics, Literature}\} \cap \{\text{Mathematics, Literature}\}.$$

In order to illustrate in a simple way the concept of rough approximation, let us constrain our analysis to the reduct {Mathematics, Literature}. Let us consider student *S*4. His positive dominance cone $D_{\{\text{Mathematics, Literature}\}}^+(S4)$ is composed of all the students having evaluations not worse than him on Mathematics and Literature, i.e., of all the students dominating him with respect to Mathematics and Literature. Thus,

$$D_{\{\text{Mathematics, Literature}\}}^+(S4) = \{S4, S5, S6\}.$$

On the other hand, the negative dominance cone of student *S*4, $D_{\{\text{Mathematics, Literature}\}}^-(S4)$, is composed of all the students having evaluations not better than him on Mathematics and Literature, i.e., of all the students dominated by him with respect to Mathematics and Literature. Thus,

$$D_{\{\text{Mathematics, Literature}\}}^-(S4) = \{S1, S2, S3, S4, S7, S8\}.$$

Similar dominance cones can be obtained for all the students from Table 6. For example, for *S*2, the dominance cones are

$$D_{\{\text{Mathematics, Literature}\}}^+(S2) = \{S1, S2, S3, S4, S5, S6\}.$$

and

$$D_{\{\text{Mathematics, Literature}\}}^-(S2) = \{S2, S7\}.$$

The rough approximations can be calculated using dominance cones. Let us consider, for example, the lower approximation of the set of students having a “good” overall evaluation $\underline{P}(CI_{\text{good}}^{\geq})$, with $P = \{\text{Mathematics, Literature}\}$. Notice that $\underline{P}(CI_{\text{good}}^{\geq}) = \{S4, S5, S6\}$, because positive dominance cones of students $S4, S5$, and $S6$ are all included in the set of students with an overall evaluation “good.” In other words, this means that there is no student dominating $S4$ or $S5$ or $S6$ while having an overall evaluation worse than “good.” From the viewpoint of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y dominates $S4$ or $S5$ or $S6$ is a *sufficient* condition to conclude that y is a “good” student.

The upper approximation of the set of students with a “good” overall evaluation is $\overline{P}(CI_{\text{good}}^{\geq}) = \{S4, S5, S6\}$, because negative dominance cones of students $S4, S5$, and $S6$ have a nonempty intersection with the set of students having a “good” overall evaluation. In other words, this means that for each one of the students $S4, S5$, and $S6$, there is at least one student dominated by him with an overall evaluation “good.” From the point of view of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y dominates $S4$ or $S5$ or $S6$ is a *possible* condition to conclude that y is a “good” student.

Let us observe that for the set of criteria $P = \{\text{Mathematics, Literature}\}$, the lower and upper approximations of the set of “good” students are the same. This means that examples of ordinal classification concerning this decision class are all consistent. This is not the case, however, for the examples concerning the union of decision classes “at least medium.” For this upward union, the rough approximations are $\underline{P}(CI_{\text{medium}}^{\geq}) = \{S3, S4, S5, S6\}$ and $\overline{P}(CI_{\text{medium}}^{\geq}) = \{S1, S2, S3, S4, S5, S6\}$. The difference between $\overline{P}(CI_{\text{medium}}^{\geq})$ and $\underline{P}(CI_{\text{medium}}^{\geq})$, i.e., the boundary $Bn_P(CI_{\text{medium}}^{\geq}) = \{S1, S2\}$, is

composed of students with inconsistent overall evaluations, which has already been noticed above. From the viewpoint of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y is dominated by $S1$ and dominates $S2$ is a condition to conclude that y can obtain an overall evaluation “at least medium” with some doubts.

Until now, rough approximations of only upward unions of decision classes have been considered. It is interesting, however, to calculate also rough approximations of downward unions of decision classes. Let us consider first the lower approximation of the set of students having “at most medium” overall evaluation $\underline{P}(CI_{\text{medium}}^{\leq})$. Observe that $\underline{P}(CI_{\text{medium}}^{\leq}) = \{S1, S2, S3, S7, S8\}$, because the negative dominance cones of students $S1, S2, S3, S7$, and $S8$ are all included in the set of students with overall evaluation “at most medium.” In other words, this means that there is no student dominated by $S1$ or $S2$ or $S3$ or $S7$ or $S8$ while having an overall evaluation better than “medium.” From the viewpoint of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y is dominated by $S1$ or $S2$ or $S3$ or $S7$ or $S8$ is a *sufficient* condition to conclude that y is an “at most medium” student.

The upper approximation of the set of students with an “at most medium” overall evaluation is $\overline{P}(CI_{\text{medium}}^{\leq}) = \{S1, S2, S3, S7, S8\}$, because the positive dominance cones of students $S1, S2, S3, S7$, and $S8$ have a nonempty intersection with the set of students having an “at most medium” overall evaluation. In other words, this means that for each one of the students $S1, S2, S3, S7$, and $S8$, there is at least one student dominating him with an overall evaluation “at most medium.” From the viewpoint of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y is dominated by $S1$ or $S2$ or $S3$ or $S7$ or $S8$ is a *possible* condition to conclude that y is an “at most medium” student.

Finally, the lower and upper approximations of the set of students having a “bad” overall evaluation are $\underline{P}(CI_{bad}^{\leq}) = \{S7, S8\}$ and $\overline{P}(CI_{bad}^{\leq}) = \{S1, S2, S7, S8\}$. The difference between $\overline{P}(CI_{bad}^{\leq})$ and $\underline{P}(CI_{bad}^{\leq})$, i.e., the boundary $Bn_P(CI_{bad}^{\leq}) = \{S1, S2\}$ is composed of students with inconsistent overall evaluations, which has already been noticed above. From the viewpoint of decision-making, this means that, taking into account the available information about evaluation of students on Mathematics and Literature, the fact that student y is dominated by $S1$ and dominates $S2$ is a condition to conclude that y can obtain an overall evaluation “bad” with some doubts. Observe, moreover, that $Bn_P(CI_{medium}^{\geq}) = Bn_P(CI_{bad}^{\leq}) = \{S1, S2\}$.

Given the above rough approximations with respect to the set of criteria $P = \{\text{Mathematics, Literature}\}$, one can induce a set of decision rules representing the preferences of the jury. The idea is that evaluation profiles of students belonging to the lower approximations can serve as a base for some certain rules, while evaluation profiles of students belonging to the boundaries can serve as a base for some approximate rules. The following decision rules have been induced (between parentheses there are id’s of students supporting the corresponding rule; the student being a rule base is underlined):

Rule (1) if the evaluation on Mathematics is (at least) good, and the evaluation on Literature is at least medium, then the overall evaluation is (at least) good, $\{\underline{S4}, S5, S6\}$.

Rule (2) if the evaluation on Mathematics is at least medium, and the evaluation on Literature is at least medium, then the overall evaluation is at least medium, $\{\underline{S3}, S4, S5, S6\}$.

Rule (3) if the evaluation on Mathematics is at least medium, and the evaluation on Literature is (at most) bad, then the overall evaluation is bad or medium, $\{\underline{S1}, S2\}$.

Rule (4) if the evaluation on Mathematics is at least medium, then the overall evaluation is at least medium, $\{\underline{S2}, S3, S4, S5, S6\}$.

Rule (5) if the evaluation on Literature is (at most) bad, then the overall evaluation is at most medium, $\{\underline{S1}, S2, S7\}$.

Rule (6) if the evaluation on Mathematics is (at most) bad, then the overall evaluation is (at most) bad, $\{\underline{S7}, S8\}$.

Notice that rules (1)–(2), (4)–(7) are certain, while rule (3) is an approximate one. These rules represent knowledge discovered from the available information. In the current context, the knowledge is interpreted as a preference model of the jury. A characteristic feature of the syntax of decision rules representing preferences is the use of expressions “at least” or “at most” a value; in case of extreme values (“good” and “bad”), these expressions are put in parentheses because there is no value above “good” and below “bad.”

Even if one can represent all the knowledge using only one reduct of the set of criteria (as it is the case using $P = \{\text{Mathematics, Literature}\}$), when considering a larger set of criteria than a reduct, one can obtain a more synthetic representation of knowledge, i.e., the number of decision rules or the number of elementary conditions, or both of them, can get smaller. For example, considering the set of all three criteria, $\{\text{Mathematics, Physics, Literature}\}$, one can induce a set of decision rules composed of the above rules (1), (2), (3), and (6), plus the following:

Rule (7) if the evaluation on Physics is at most medium, and the evaluation on Literature is at most medium, then the overall evaluation is at most medium, $\{S1, S2, \underline{S3}, S7, S8\}$.

Thus, the complete set of decision rules induced from Table 3 is composed of 5 instead of 6 rules.

Once accepted by the DM, these rules represent his/her preference model. Assuming that rules (1)–(7) in our example represent the preference model of the jury, it can be used to evaluate new students. For example, student $S9$ who is “medium” in Mathematics and Physics, and “good” in Literature would be evaluated as “medium,” because his profile matches the premise of rule (2), having as consequence an overall evaluation at least “medium.” The overall evaluation of $S9$ cannot be “good,” because his profile does not match any rule having as consequence an overall

evaluation “good” (in the considered example, the only rule of this type is rule (1) whose premise is not matched by the profile of S_9).

Robust Ordinal Regression for DRSA

Taking into account that several complete and minimal sets of decision rules can be induced from rough approximation of a decision table, and that each such set of decision rules can propose a different ordinal classification of objects not present in the decision table, it is natural to search for a methodology able to consider this plurality of preference representation by sets of decision rules. Similar concerns arise in multiple criteria decision aiding based on other preference models, like value functions or outranking relations. To cope with the problem of non-univocal preference representation by models induced from sets of decision examples, a general methodology has been proposed, called Robust Ordinal Regression (ROR) (see Greco et al. (2008b) for a seminal paper on this topic, and Corrente et al. (2013) for an updated survey). ROR takes into account all instances of an applied preference model (a value function, or an outranking relation, or a minimal set of decision rules) which are compatible with the preference information supplied by the DM, i.e., all instances of preference model parameters that reproduce the DM's preference statements expressing his/her value system. The basic idea of ROR is the distinction between necessary and possible relation resulting from application of all compatible instances of a preference model on the considered set of objects. For example, object a is necessarily preferred to another object b if this is true for all compatible instances of the adopted preference model, while a is possibly preferred to b if this is true for at least one compatible instance. ROR has been applied to DRSA in the following procedure (Kadziński et al. 2014). First, the whole family of minimal sets of minimal decision rules is induced. Then, following an idea proposed in Greco et al. (2010b) for ordinal classification based on value functions, an object x is *necessarily assigned* to class Cl_h , if and only if it is assigned to Cl_h by all sets of minimal rules, and x is *possibly assigned* to Cl_h , if and only if x is assigned to Cl_h by at least one minimal set of rules. Moreover, it is possible to compute the

class acceptability indices defined as the share of compatible sets of rules assigning an object x to a single class or to a set of contiguous classes. The class acceptability indices can be interpreted as a probability that object x is assigned to the considered class or set of contiguous classes.

DRSA on a Pairwise Comparison Table for Multiple Criteria Choice and Ranking Problems

Multiple Criteria Choice and Ranking Problems

Ordinal classification decisions are based on absolute evaluation of objects on multiple criteria; however, multiple criteria choice and ranking decisions are based on preference relations between objects. Decision table including examples of ordinal classification does not contain information about preference relations between objects, thus, in order to apply DRSA to multiple criteria choice and ranking problems, a different representation of the input preference information is needed.

To handle binary relations within the rough set approach, it has been proposed in Greco et al. (1999a, b) to operate on, so-called, *pairwise comparison table* (PCT), i.e., a decision table including pairs of objects for which multiple criteria evaluations and a comprehensive preference relation are known. PCT represents preference information provided by the DM in the form of decision examples (pairwise comparisons of objects).

Similarly to ordinal classification, decision examples concerning multiple criteria choice and ranking may be inconsistent with respect to the dominance principle; however, interpretation of the inconsistency is different: it occurs when preferences of a pair of objects, say (a, b) , on all considered criteria are not weaker than preferences of another pair of objects, say (c, d) , on these criteria; however, the comprehensive preference of object a over object b is weaker than the comprehensive preference of object c over object d .

The Pairwise Comparison Table

Similarly to the ordinal classification, let us consider a finite set of criteria $F = \{f_1, \dots, f_n\}$, the set of their indices $I = \{1, \dots, n\}$, and a finite

universe of objects (actions, solutions, alternatives) U . For any criterion $i \in I$, let T_i be a finite set of binary relations defined on U on the basis of the evaluations of objects from U with respect to the considered criterion i , such that $\forall(x, y) \in U \times U$ exactly one binary relation $t \in T_i$ is verified; $t \in T_i$ has the meaning of a *preference relation* for a pair of objects on a particular criterion i . More precisely, given value set V_i , $i \in I$, if $v'_i, v''_i \in V_i$ are the respective evaluations of $x, y \in U$ on criterion i , and $(x, y) \in t$, with $t \in T_i$, then for each $w, z \in U$ having the same evaluations v'_i, v''_i on i , $(w, z) \in t$. For interesting applications, it should be $|T_i| \geq 2$, $\forall i \in I$. Furthermore, let T_d be a set of binary relations defined on set U , such that at most one binary relation $t \in T_d$ is verified $\forall(x, y) \in U \times U$; $t \in T_d$ has the meaning of a *comprehensive preference relation* for a pair of objects (comprehensive pairwise comparison).

The preference information provided by the DM has the form of pairwise comparisons of some reference objects from $B \subseteq U$. These decision examples are presented in the pairwise comparison table (PCT), defined as information table $S_{PCT} = \langle \widehat{B}, F \cup \{d\}, T_F \cup T_{d,g} \rangle$, where $\widehat{B} \subseteq B \times B$ is a non-empty set of *exemplary pairwise comparisons of reference objects*, $T_F = \bigcup_{i \in I} T_i$, d is a decision corresponding to the comprehensive pairwise comparison (comprehensive preference relation), and $g : \widehat{B} \times (F \cup \{d\}) \rightarrow T_F \cup T_d$ is a total function such that $g[(x, y), i] \in T_i$, $\forall(x, y) \in U \times U$ and $\forall i \in I$, and $g[(x, y), d] \in T_d$, $\forall(x, y) \in \widehat{B}$. It follows that for any pair of reference objects $(x, y) \in \widehat{B}$, there is verified one and only one binary relation $t \in T_d$. Thus, T_d induces a partition of $\widehat{B}$. In fact, information Table S_{PCT} can be seen as decision table, since the set of considered criteria F and decision d are distinguished.

It is assumed that the exemplary pairwise comparisons provided by the DM can be represented in terms of *graded preference relations* (for example, “very weak preference,” “weak preference,” “strict preference,” “strong preference,” and “very strong preference”) P_i^h : $\forall i \in I$ and $\forall(x, y) \in U \times U$,

$$T_i = \{P_i^h, h \in H_i\},$$

where H_i is a particular subset of the relative integers and

$xP_i^h y, h > 0$, means that object x is preferred to object y by degree h with respect to criterion i ,

$xP_i^h y, h < 0$, means that object x is not preferred to object y by degree h with respect to criterion i ,
 $xP_i^0 y$ means that object x is similar (asymmetrically indifferent) to object y with respect to criterion i .

Of course, $\forall i \in I$ and $\forall(x, y) \in U \times U$, it holds:

$$[xP_i^h y, h \geq 0] \Leftrightarrow [yP_i^k x, k \leq 0].$$

The set of binary relations T_d may be defined in a similar way, but $xP_d^h y$ means that object x is comprehensively preferred to object y by degree h .

Technically, the modeling of the binary relation P_i^h , i.e., the assessment of h , can be organized as follows:

- First, it is observed that criterion i is a function $f_i : U \rightarrow \mathfrak{R}$ increasing with respect to the preferences on i , for each $i = 1, \dots, n$.
- Then, for each $i = 1, \dots, n$, it is possible to define a function $k_i : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ which measures the *strength of the preference* (positive or negative) of x over y (e.g., $k_i[f_i(x), f_i(y)] = f_i(x) - f_i(y)$); it should satisfy the following properties $\forall x, y, z \in U$:
 - i) $f_i(x) > f_i(y) \Leftrightarrow k_i[f_i(x), f_i(z)] > k_i[f_i(y), f_i(z)]$.
 - ii) $f_i(x) > f_i(y) \Leftrightarrow k_i[f_i(z), f_i(y)] > k_i[f_i(z), f_i(x)]$.
 - iii) $f_i(x) = f_i(y) \Leftrightarrow k_i[f_i(x), f_i(y)] = 0$.
- Next, the domain of k_i can be divided into intervals, using a suitable set of thresholds Δ_i , $\forall i \in I$; these intervals are numbered in such a way that $k_i[f_i(x), f_i(y)] = 0$ belongs to interval number 0.
- The value of h in the relation $xP_i^h y$ is then equal to the number of the interval including $k_i[f_i(x), f_i(y)]$, for any $(x, y) \in U \times U$.

Actually, property (iii) can be relaxed in order to obtain a more general preference model which, for instance, does not satisfy preferential independence.

To simplify the presentation, let us consider a PCT where the set T_d is composed of two binary relations defined on U :

- x outranks y (denotation xSy or $(x, y) \in S$), where $(x, y) \in \widehat{B}$.
- x does not outrank y (denotation $xS^c y$ or $(x, y) \in S^c$), where $(x, y) \in \widehat{B}$.

and $S \cup S^c = \widehat{B}$, where “ x outranks y ” means “ x is at least as good as y ”; observe that the binary relation S is reflexive, but neither necessarily transitive nor complete. In Fortemps et al. (2008), a more general PCT was considered, where the set T_d is composed of multi-graded binary relations defined on U .

Approximation by Means of Graded Dominance Relations

Let $H_P = \bigcap_{i \in P} H_i$, $\forall P \subseteq I$. Given $P \subseteq I$ and $h \in H_P$, $\forall (x, y) \in U \times U$ it is said that x positively dominates y by degree h with respect to criteria from P iff $xP_i^{c_i} y$ with $c_i \geq h$, $\forall i \in P$. Analogously, $\forall (x, y) \in U \times U$, x negatively dominates y by degree h with respect to criteria from P iff $xP_i^{c_i} y$ with $c_i \leq h$, $\forall i \in P$. Therefore, each $P \subseteq I$ and $h \in H_P$ generate two binary relations (possibly empty) on U , called positive P -dominance by degree h (denotation D_{+P}^h) and negative P -dominance by degree h (denotation D_{-P}^h), respectively. They satisfy the following conditions:

- (P1) if $(x, y) \in D_{+P}^h$, then $(x, y) \in D_{+R}^k$
for each $R \subseteq P$ and $k \leq h$.
- (P2) if $(x, y) \in D_{-P}^h$, then $(x, y) \in D_{-R}^k$
for each $R \subseteq P$ and $k \geq h$.

In Greco et al. (1999b), it has been proposed to approximate the outranking relation S by means of the dominance relation D_{+P}^h . Therefore, S is considered a rough binary relation.

The P -lower approximation of S (denotation $\underline{P}(S)$) and the P -upper approximation of S (denotation $\overline{P}(S)$) are defined, respectively, as:

$$\underline{P}(S) = \bigcup_{h \in H_P} \left\{ \left(D_{+P}^h \cap \widehat{B} \right) \subseteq S \right\},$$

$$\overline{P}(S) = \bigcap_{h \in H_P} \left\{ \left(D_{+P}^h \cap \widehat{B} \right) \supseteq S \right\}.$$

Remembering (P1), $\underline{P}(S)$ may be interpreted as the dominance relation D_{+P}^h having the largest intersection with $\widehat{B}$ included in the outranking relation S , and $\overline{P}(S)$ as the dominance relation D_{+P}^h including S and having the smallest intersection with $\widehat{B}$.

Analogously, it is possible to approximate the relation S^c by means of the dominance relation D_{-P}^h . Observe that, in general, the definitions of the approximations of S and S^c do not satisfy the condition of complementarity, i.e., it is not true, in general, that $\underline{P}(S)$ is equal to $\widehat{B} - \overline{P}(S^c)$ and that $\underline{P}(S^c)$ is equal to $\widehat{B} - \overline{P}(S)$. This is because S and S^c are approximated using two different relations, D_{+P}^h and D_{-P}^h , respectively. Nevertheless, the approximations thus obtained constitute a good basis for the generation of simple decision rules.

Decision Rules

It is possible to represent preferences of the DM revealed in terms of exemplary pairwise comparisons contained in a given PCT, using decision rules. Since approximations of S and S^c were made using graded dominance relations, it is possible to induce decision rules being propositions of the following type:

- D_{++} -decision rule: if $x D_{+P}^h y$, then xSy .
- D_{+-} -decision rule: if not $x D_{+P}^h y$, then $xS^c y$.
- D_{-+} -decision rule: if $x D_{-P}^h y$, then xSy .
- D_{--} -decision rule: if not $x D_{-P}^h y$, then $xS^c y$.

where P is a non-empty subset of I . Therefore, for example, a D_{++} -decision rule is a proposition of the type: “if x positively dominates y by degree h with respect to criteria from P , then x outranks y .”

A constructive definition of these rules may be given, being a kind of implication supported by the existence of at least one pair of objects from $\widehat{B}$ satisfying one of the four propositions listed above, and by the absence of pairs from $\widehat{B}$ contradicting it. Thus, for example, if

- There exists at least one pair $(w, z) \in \widehat{B}$ such that $wD_{+P}^h z$ and $wS^c z$ and
- There does not exist any pair $(v, u) \in \widehat{B}$ such that $vD_{+P}^h u$ and $vS^c u$,
- Then “if $xD_{+P}^h y$, then xSy ” is accepted as a D_{++-} decision rule.

A D_{++-} decision rule “if $xD_{+P}^h y$, then xSy ” is said to be *minimal* if there does not exist any rule “if $xD_{+R}^k y$, then xSy ” such that $R \subseteq P$ and $k \leq h$. Analogous definitions hold for the other cases. In other words, a minimal decision rule is a kind of implication for which there is no other implication whose premise is of at least the same weakness and whose consequence is of at least the same strength.

The following results show connections of the decision rules with the P -lower and P -upper approximations of S and S^c (Greco et al. 1999b):

- D_{++-} minimal decision rule “if $xD_{+P}^h y$, then xSy ” is supported by pairs of objects belonging to $\underline{P}(S) = D_{+P}^h \cap \widehat{B}$.
- D_{--} minimal decision rule “if $xD_{-P}^h y$, then $xS^c y$ ” is supported by pairs of objects belonging to $\underline{P}(S^c) = D_{-P}^h \cap \widehat{B}$.
- D_{+-} minimal decision rule “if not $xD_{+P}^h y$, then $xS^c y$ ” is supported by pairs of objects belonging to $\overline{P}(S) = D_{+P}^h \cap \widehat{B}$.
- D_{-+} minimal decision rule “if not $xD_{-P}^h y$, then xSy ” is supported by pairs of objects belonging to $\overline{P}(S^c) = D_{-P}^h \cap \widehat{B}$.

Application of the Decision Rules and Final Recommendation

In order to obtain a recommendation in the multiple criteria choice or ranking problems with respect to a set of objects $M \subseteq U$, the decision rules induced from the approximations of S and S^c

(defined with respect to reference objects from B) should be applied on set $M \times M$. The application of the rules to any pair of objects $(u, v) \in M \times M$ establishes the presence (uSv) or the absence ($uS^c v$) of outranking with respect to (u, v) . More precisely,

- From D_{++-} decision rule “if $xD_{+P}^h y$ then xSy ” and from $uD_{+P}^h v$, one concludes uSv .
- From D_{+-} decision rule “if not $xD_{+P}^h y$ then $xS^c y$ ” and from not $uD_{+P}^h v$, one concludes $uS^c v$.
- From D_{-+} decision rule “if not $xD_{-P}^h y$, then xSy ” and from not $uD_{-P}^h v$, one concludes uSv .
- From D_{--} decision rule “if $xD_{-P}^h y$, then $xS^c y$ ” and from $uD_{-P}^h v$, one concludes $uS^c v$.

After the application of the decision rules to each pair of objects $(u, v) \in M \times M$, one of the following four situations may occur:

- uSv and not $uS^c v$, that is *true* outranking (denotation $uS^T v$).
- $uS^c v$ and not uSv , that is *false* outranking (denotation $uS^F v$).
- uSv and $uS^c v$, that is *contradictory* outranking (denotation $uS^K v$).
- not uSv and not $uS^c v$, that is *unknown* outranking (denotation $uS^U v$).

The four above situations, which together constitute the so-called four-valued outranking (see Tsoukias and Vincke 1995), have been introduced to underline the *presence* and the *absence* of *positive* and *negative* reasons for the outranking. Moreover, they make it possible to distinguish contradictory situations from unknown ones.

The following theorem underlines the operational importance of the minimal decision rules (Greco et al. 1999b): the application of *all* the decision rules obtained for a given S_{PCT} to a pair of objects $(u, v) \in M \times M$ results in the same outranking relations S and S^c as those obtained from the application of the *minimal* decision rules *only*. Therefore, the set of the minimal decision rules totally characterizes the preferences of the DM contained in S_{PCT} .

A final *recommendation* can be obtained upon a suitable exploitation of the presence and the absence of outranking S and S^c on M . A possible exploitation procedure consists in calculating a specific score, called *Net Flow Score*, for each object $x \in M$:

$$S_{nf}(x) = S^{++}(x) - S^{+-}(x) + S^{-+}(x) - S^{--}(x),$$

where

$$S^{++}(x) = |\{y \in M: \text{there is at least one decision rule which affirms } xSy\}|.$$

$$S^{+-}(x) = |\{y \in M: \text{there is at least one decision rule which affirms } ySx\}|.$$

$$S^{-+}(x) = |\{y \in M: \text{there is at least one decision rule which affirms } yS^c x\}|.$$

$$S^{--}(x) = |\{y \in M: \text{there is at least one decision rule which affirms } xS^c y\}|.$$

The recommendation in multiple criteria ranking problems consists of the total preorder determined by $S_{nf}(x)$ on M ; in multiple criteria choice problems, it consists of the object(s) $x^* \in M$ such that $S_{nf}(x^*) = \max_{x \in M} S_{nf}(x)$.

The procedure described above has been characterized with reference to a number of desirable properties in Greco et al. (1998b).

Approximation by Means of Multi-graded Dominance Relations

The graded dominance relation introduced above assumes a common grade of preference for all the considered criteria. While this permits a simple calculation of the approximations and of the resulting decision rules, it is lacking in precision. A dominance relation allowing a different degree of preference for each considered criterion (multi-graded dominance) gives a far more accurate picture of the preference information contained in the pairwise comparison table $\mathcal{S}_{PCT}$ (Greco et al. 1999a, 2000, 2001a).

More formally, given $P \subseteq I$ ($P \neq \emptyset$), (x, y) , $(w, z) \in U \times U$, (x, y) is said to *dominate* (w, z) with respect to criteria from P (denotation $(x, y)D_P(w, z)$) if x is preferred to y at least as strongly as w is preferred to z with respect to

each $i \in P$. Precisely, “at least as strongly as” means “by at least the same degree,” i.e., $hi \geq ki$, where $hi, ki \in H_i$, $xP_i^{hi}y$, and $wP_i^{ki}z$, $\forall i \in P$. Let $D_{\{i\}}$ be the dominance relation confined to the single criterion $i \in P$. The binary relation $D_{\{i\}}$ is reflexive $((x, y)D_{\{i\}}(x, y), \forall (x, y) \in U \times U)$, transitive $((x, y)D_{\{i\}}(w, z) \text{ and } (w, z)D_{\{i\}}(u, v) \text{ imply } (x, y)D_{\{i\}}(u, v), \forall (x, y), (w, z), (u, v) \in U \times U)$, and complete $((x, y)D_{\{i\}}(w, z) \text{ and/or } (w, z)D_{\{i\}}(x, y), \forall (x, y), (w, z) \in U \times U)$. Therefore, $D_{\{i\}}$ is a complete preorder on $U \times U$. Since the intersection of complete preorders is a partial preorder and $D_P = \bigcap_{i \in P} D_{\{i\}}$, $P \subseteq I$, then the dominance relation D_P is a partial preorder on $U \times U$.

Let $R \subseteq P \subseteq I$ and $(x, y), (u, v) \in U \times U$; then the following implication holds:

$$(x, y)D_P(u, v) \Rightarrow (x, y)D_R(u, v).$$

Given $P \subseteq I$ and $(x, y) \in U \times U$, let us introduce the positive dominance set (denotation $D_P^+(x, y)$) and the negative dominance set (denotation $D_P^-(x, y)$):

$$D_P^+(x, y) = \{(w, z) \in U \times U : (w, z)D_P(x, y)\},$$

$$D_P^-(x, y) = \{(w, z) \in U \times U : (x, y)D_P(w, z)\}.$$

Using the dominance relation D_P , it is possible to define P -lower and P -upper approximations of the outranking relation S with respect to $P \subseteq I$, respectively, as:

$$\underline{P}(S) = \{(x, y) \in \widehat{B} : D_P^+(x, y) \subseteq S\},$$

$$\overline{P}(S) = \bigcup_{(x, y) \in S} D_P^+(x, y).$$

Analogously, it is possible to define the approximations of S^c :

$$\underline{P}(S^c) = \{(x, y) \in \widehat{B} : D_P^-(x, y) \subseteq S^c\},$$

$$\overline{P}(S^c) = \bigcup_{(x, y) \in S^c} D_P^-(x, y).$$

It may be proved that

$$\begin{aligned}\underline{P}(S) &\subseteq S \subseteq \overline{P}(S) \\ \underline{P}(S^c) &\subseteq S^c \subseteq \overline{P}(S^c).\end{aligned}$$

Furthermore, the following complementarity properties hold:

$$\begin{aligned}\underline{P}(S) &= \widehat{B} - \overline{P}(S^c), & \overline{P}(S) &= \widehat{B} - \underline{P}(S^c), \\ \underline{P}(S^c) &= \widehat{B} - \overline{P}(S), & \overline{P}(S^c) &= \widehat{B} - \underline{P}(S).\end{aligned}$$

The P -boundaries (P -doubtful regions) of S and S^c are defined as

$$\begin{aligned}Bn_P(S) &= \overline{P}(S) - \underline{P}(S), \\ Bn_P(S^c) &= \overline{P}(S^c) - \underline{P}(S^c).\end{aligned}$$

It is easy to prove that $Bn_P(S) = Bn_P(S^c)$.

The concepts of quality of approximation, reducts, and core can be extended also to the approximation of the outranking relation by multi-graded dominance relations. In particular,

$$\gamma_P = \frac{|\underline{P}(S) \cup \underline{P}(S^c)|}{|\widehat{B}|}$$

defines the *quality of approximation* of S and S^c by $P \subseteq I$. It expresses the ratio of all pairs of objects $(x, y) \in \widehat{B}$ correctly assigned to S and S^c by the set P of criteria, to all the pairs of objects contained in $\widehat{B}$. Each minimal subset $P' \subseteq P$ such that $\gamma_{P'} = \gamma_P$ is called a *reduct* of P (denotation $RED_{PCT}(P)$). Let us remark that S_{PCT} can have more than one reduct. The intersection of all reducts is called the *core* (denotation $CORE_{PCT}(P)$).

Using the approximations defined above, it is then possible to induce a generalized description of the preference information contained in a given S_{PCT} in terms of suitable decision rules. The syntax of these rules is based on the concept of upward cumulated preferences (denotation $P_i^{\geq h}$) and downward cumulated preferences (denotation $P_i^{\leq h}$), having the following interpretation:

- $xP_i^{\geq h}y$ means “ x is preferred to y with respect to i by at least degree h .”
- $xP_i^{\leq h}y$ means “ x is preferred to y with respect to i by at most degree h .”

Exact definition of the cumulated preferences, for each $(x, y) \in U \times U$, $i \in I$ and $h \in H_i$, is the following:

- $xP_i^{\geq h}y$ if $xP_i^k y$, where $k \in H_i$ and $k \geq h$.
- $xP_i^{\leq h}y$ if $xP_i^k y$, where $k \in H_i$ and $k \leq h$.

Using the above concepts, three types of decision rules can be obtained:

1. $D_{\geq-}$ decision rules, being statements of the type:
if $xP_{i1}^{\geq h(i1)}y$ and $xP_{i2}^{\geq h(i2)}y$ and ... $xP_{ip}^{\geq h(ip)}y$,
then xSy ,
where $P = \{i1, \dots, ip\} \subseteq I$ and $(h(i1), \dots, h(ip)) \in H_{i1} \times \dots \times H_{ip}$; these rules are supported by pairs of objects from the P -lower approximation of S only.
2. $D_{\leq-}$ decision rules, being statements of the type:
if $xP_{i1}^{\leq h(i1)}y$ and $xP_{i2}^{\leq h(i2)}y$ and ... $xP_{ip}^{\leq h(ip)}y$,
then $xS^c y$,
where $P = \{i1, \dots, ip\} \subseteq I$ and $(h(i1), \dots, h(ip)) \in H_{i1} \times \dots \times H_{ip}$; these rules are supported by pairs of objects from the P -lower approximation of S^c only.
3. $D_{\geq \leq-}$ decision rules, being statements of the type:
if $xP_{i1}^{\geq h(i1)}y$ and $xP_{i2}^{\geq h(i2)}y$ and ... $xP_{ik}^{\geq h(ik)}y$
and ... $xP_{ik+1}^{\leq h(ik+1)}y$ and ... $xP_{ip}^{\leq h(ip)}y$, then xSy
or $xS^c y$,
where $O' = \{i1, \dots, ik\} \subseteq I$, $O'' = \{ik+1, \dots, ip\} \subseteq I$, $P = O' \cup O''$, O' and O'' not necessarily disjoint, $(h(i1), \dots, h(ip)) \in H_{i1} \times H_{i2} \times \dots \times H_{ip}$; these rules are supported by objects from the P -boundary of S and S^c only.

Dominance Without Degrees of Preference

The degree of graded preference considered in subsection “[The Pairwise Comparison Table](#)” is defined on a *quantitative* scale of the strength of preference k_i , $i \in I$. However, in many real world problems, the existence of such a quantitative scale is rather questionable. Roy (1999) distinguishes the following cases:

- Preferences expressed on an *ordinal* scale: this is the case where the difference between two evaluations has no clear meaning.
- Preferences expressed on a *quantitative* scale: this is the case where the scale is defined with reference to a unit clearly identified, such that it is meaningful to consider an origin (zero) of the scale and ratios between evaluations (ratio scale).
- Preferences expressed on a *numerical non-quantitative* scale: this is an intermediate case between the previous two; there are two well-known particular cases:
 - *Interval scale*, where it is meaningful to compare ratios between differences of pairs of evaluations.
 - Scale for which a *complete preorder* can be defined on all possible pairs of evaluations.

The preference scale has also been considered within economic theory (e.g., Shoemaker 1982), where cardinal utility is distinguished from ordinal utility: the former deals with a strength of preference, while, for the latter, this concept is meaningless. From this point of view, preferences expressed on an ordinal scale refer to *ordinal utility*, while preferences expressed on a quantitative scale or a numerical nonquantitative scale deal with *cardinal utility*.

The strength of preference k_i and, therefore, the graded preference considered in subsection “[The Pairwise Comparison Table](#)” is meaningful when the scale is quantitative or numerical non-quantitative. If the information about k_i is non-available, then it is possible to define a rough approximation of S and S^c using a specific dominance relation between pairs of objects from $U \times U$, defined on an ordinal scale represented by evaluations $f_i(x)$ on criterion i , for $x \in U$ (Greco et al. 1999a). Let us explain this latter case in more details.

Let I^O be the set of criteria expressing preferences on an ordinal scale, and I^N , the set of criteria expressing preferences on a quantitative scale or a numerical nonquantitative scale, such that $I^O \cup I^N = I$ and $I^O \cap I^N = \emptyset$. Moreover, for each $P \subseteq I$, P^O denotes the subset of P composed

of criteria expressing preferences on an ordinal scale, i.e., $P^O = P \cap I^O$, and P^N the subset of P composed of criteria expressing preferences on a quantitative scale or a numerical non-quantitative scale, i.e., $P^N = P \cap I^N$. Of course, for each $P \subseteq I$, $P = P^N \cup P^O$ and $P^O \cap P^N = \emptyset$.

If $P = P^N$ and $P^O = \emptyset$, then the definition of dominance is the same as in the case of multi-graded dominance (subsection “[Approximation by Means of Multi-Graded Dominance Relations](#)”). If $P = P^O$ and $P^N = \emptyset$, then, given $(x, y), (w, z) \in U \times U$, the pair (x, y) is said to dominate the pair (w, z) with respect to P if, for each $i \in P$, $f_i(x) \geq f_i(w)$ and $f_i(z) \geq f_i(y)$. Let $D_{\{i\}}$ be the dominance relation confined to the single criterion $i \in P^O$. The binary relation $D_{\{i\}}$ is reflexive $((x, y)D_{\{i\}}(x, y), \forall (x, y) \in U \times U)$, transitive $((x, y)D_{\{i\}}(w, z) \text{ and } (w, z)D_{\{i\}}(u, v) \text{ imply } (x, y)D_{\{i\}}(u, v), \forall (x, y), (w, z), (u, v) \in U \times U)$, but non-complete (it is possible that *not* $(x, y)D_{\{i\}}(w, z)$ and *not* $(w, z)D_{\{i\}}(x, y)$ for some $(x, y), (w, z) \in U \times U$). Therefore, $D_{\{i\}}$ is a partial preorder. Since the intersection of partial preorders is also a partial preorder and $D_P = \bigcap_{i \in P} D_{\{i\}}$,

$P = P^O$, then the dominance relation D_P is also a partial preorder.

If some criteria from $P \subseteq I$ express preferences on a quantitative or a numerical nonquantitative scale and others on an ordinal scale, i.e., if $P^N \neq \emptyset$ and $P^O \neq \emptyset$, then, given $(x, y), (w, z) \in U \times U$, the pair (x, y) is said to dominate the pair (w, z) with respect to criteria from P , if (x, y) dominates (w, z) with respect to both P^N and P^O . Since the dominance relation with respect to P^N is a partial preorder on $U \times U$ (because it is a multi-graded dominance) and the dominance with respect to P^O is also a partial preorder on $U \times U$ (as explained above), then also the dominance D_P , being the intersection of these two dominance relations, is a partial preorder. In consequence, all the concepts introduced in the previous subsection can be restored using this specific definition of dominance relation.

Using the approximations of S and S^c based on the dominance relation defined above, it is possible to induce a generalized description of the available preference information in terms of

decision rules. These decision rules are of the same type as the rules already introduced in the previous subsection; however, the conditions on criteria from I^O are expressed directly in terms of evaluations belonging to value sets of these criteria. Let $F_i = \{f_i(x), x \in U\}$, $i \in I^O$. The decision rules have in this case the following syntax:

1. $D_{\geq-}$ decision rule, being a statement of the type:

if $xP_{il}^{\geq h(i1)}y$ and ... $xP_{ie}^{\geq h(ie)}y$ and $f_{ie+1}(x) \geq r_{ie+1}$ and $f_{ie+1}(y) \leq s_{ie+1}$ and ... $f_{ip}(x) \geq r_{ip}$ and $f_{ip}(y) \leq s_{ip}$, then xSy ,
where $P = \{i1, \dots, ip\} \subseteq I$, $P^N = \{i1, \dots, ie\}$, $P^O = \{ie + 1, \dots, ip\}$, $(h(i1), \dots, h(ie)) \in H_{i1} \times \dots \times H_{ie}$ and $(r_{ie+1}, \dots, r_{ip}), (s_{ie+1}, \dots, s_{ip}) \in F_{ie+1} \times \dots \times F_{ip}$; these rules are supported by pairs of objects from the P -lower approximation of S only.

2. $D_{\leq-}$ decision rule, being a statement of the type:

if $xP_{il}^{\leq h(i1)}y$ and ... $xP_{ip}^{\leq h(ip)}y$ and $f_{ie+1}(x) \leq r_{ie+1}$ and $f_{ie+1}(y) \geq s_{ie+1}$ and ... $f_{ip}(x) \leq r_{ip}$ and $f_{ip}(y) \geq s_{ip}$, then $xS^c y$,
where $P = \{i1, \dots, ip\} \subseteq I$, $P^N = \{i1, \dots, ie\}$, $P^O = \{ie + 1, \dots, ip\}$, $(h(i1), \dots, h(ie)) \in H_{i1} \times \dots \times H_{ie}$ and $(r_{ie+1}, \dots, r_{ip}), (s_{ie+1}, \dots, s_{ip}) \in F_{ie+1} \times \dots \times F_{ip}$; these rules are supported by pairs of objects from the P -lower approximation of S^c only.

3. $D_{\geq \leq-}$ decision rule, being a statement of the type:

if $xP_{il}^{\geq h(i1)}y$ and ... $xP_{ie}^{\geq h(ie)}y$ and $xP_{ie+1}^{\geq h(ie+1)}y$... $xP_{if}^{\leq h(if)}y$ and $f_{if+1}(x) \geq r_{if+1}$ and $f_{if+1}(y) \leq s_{if+1}$ and ... $f_{ig}(x) \geq r_{ig}$ and $f_{ig}(y) \leq s_{ig}$ and $f_{ig+1}(x) \leq r_{ig+1}$ and $f_{ig+1}(y) \geq s_{ig+1}$ and ... $f_{ip}(x) \leq r_{ip}$ and $f_{ip}(y) \geq s_{ip}$, then xSy or $xS^c y$,
where $O' = \{i1, \dots, ie\} \subseteq I$, $O'' = \{ie + 1, \dots, if\} \subseteq I$, $P^N = O' \cup O''$, O' and O'' not necessarily disjoint, $P^O = \{if + 1, \dots, ip\}$, $(h(i1), \dots, h(if)) \in H_{i1} \times \dots \times H_{if}$ and $(r_{if+1}, \dots, r_{ip}), (s_{if+1}, \dots, s_{ip}) \in F_{if+1} \times \dots \times F_{ip}$; these rules are supported by pairs of objects from the P -boundary of S and S^c only.

Various procedures for construction of a ranking recommendation using sets of the above decision rules have been characterized with respect to a set of desirable properties in Szelaq et al. (2014). Moreover, the Robust Ordinal Regression (ROR) methodology described in subsection “Robust Ordinal Regression for DRSA” has been applied to multiple-criteria ranking and choice with all compatible minimal sets of decision rules in Kadziński et al. (2015).

Example Illustrating DRSA in the Context of Multiple Criteria Choice and Ranking

The following example illustrates DRSA in the context of multiple criteria choice and ranking. Six warehouses have been described by means of four criteria:

- f_1 , capacity of the sales staff
- f_2 , perceived quality of goods
- f_3 , high traffic location
- f_4 , warehouse profit or loss

The components of the information table S are: $U = \{1, 2, 3, 4, 5, 6\}$, $F = \{f_1, f_2, f_3, f_4\}$, $I = \{1, 2, 3, 4\}$, $F_1 = \{\text{high, medium, low}\}$, $F_2 = \{\text{good, medium}\}$, $F_3 = \{\text{no, yes}\}$, $F_4 = \{\text{profit, loss}\}$, the criterion $f_i(x)$, taking values $f_1(1) = \text{high}$, $f_2(1) = \text{good}$, and so on (Table 7).

It is assumed that the DM accepts to express preferences with respect to criteria f_1, f_2, f_3 on a *numerical nonquantitative scale*, for which a complete preorder can be defined on all possible pairs of evaluations. According to this assumption, in order to build the PCT, as described in subsection “The Pairwise Comparison Table,” the DM specifies sets of possible degrees of

Rough Sets in Decision-Making, Table 7 Information table of the illustrative example

Warehouse	f_1	f_2	f_3	f_4
1	High	Good	No	Profit
2	Medium	Medium	No	Loss
3	Medium	Medium	No	Profit
4	Low	Medium	No	Loss
5	Medium	Good	Yes	Loss
6	High	Medium	Yes	Profit

preference; for example, $H_1 = \{-2, -1, 0, 1, 2\}$, $H_2 = \{-1, 0, 1\}$, $H_3 = \{-1, 0, 1\}$. Therefore, with respect to f_1 , there are the following preference relations P_1^h :

- xP_1^2y (and $yP_1^{-2}x$), meaning that x is preferred to y with respect to f_1 , if $f_1(x) = \text{high}$ and $f_1(y) = \text{low}$.
- xP_1^1y (and $yP_1^{-1}x$), meaning that x is weakly preferred to y with respect to f_1 , if $f_1(x) = \text{high}$ and $f_1(y) = \text{medium}$, or $f_1(x) = \text{medium}$ and $f_1(y) = \text{low}$.
- xP_1^0y (and yP_1^0x), meaning that x is indifferent to y with respect to f_1 , if $f_1(x) = f_1(y)$.

Analogously, with respect to f_2 and f_3 , there are the following preference relations P_2^h and P_3^h :

- xP_2^1y (and $yP_2^{-1}x$), meaning that x is weakly preferred to y with respect to f_2 , if $f_2(x) = \text{good}$ and $f_2(y) = \text{medium}$.
- xP_2^0y (and yP_2^0x), meaning that x is indifferent to y with respect to f_2 , if $f_2(x) = f_2(y)$.
- xP_3^1y (and $yP_3^{-1}x$), meaning that x is weakly preferred to y with respect to f_3 , if $f_3(x) = \text{yes}$ and $f_3(y) = \text{no}$.
- xP_3^0y (and yP_3^0x), meaning that x is indifferent to y with respect to f_3 , if $f_3(x) = f_3(y)$.

As to the comprehensive preference relation, the DM considers that, given two different warehouses $x, y \in U = \{1, 2, 3, 4, 5, 6\}$, if x makes profit and y makes loss, then xS_y and yS^c_x . Moreover, the DM accepts xS_x for each warehouse x . As to warehouses x and y which both make profit or both make loss, the DM abstains from judging whether xS_y or xS^c_y . Therefore, the set of exemplary pairwise comparisons supplied by the DM is $\hat{B} = \{(1, 1), (1, 2), (1, 4), (1, 5), (2, 1), (2, 2), (2, 3), (2, 6), (3, 2), (3, 3), (3, 4), (3, 5), (4, 1), (4, 3), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (5, 6), (6, 2), (6, 4), (6, 5), (6, 6)\}$.

At this stage, the PCT can be built as shown in Table 8.

Rough Sets in Decision-Making, Table 8 Pairwise comparison table

Pairs	P_1^h	P_2^h	P_3^h	Outranking
(1,1)	0	0	0	S
(1,2)	1	1	0	S
(1,4)	2	1	0	S
(1,5)	1	0	-1	S
(2,1)	-1	-1	0	S^c
(2,2)	0	0	0	S
(2,3)	0	0	0	S^c
(2,6)	-1		1	S^c
(3,2)	0	0	0	S
(3,3)	0	0	0	S
(3,4)	1	0	0	S
(3,5)	0	-1	-1	S
(4,1)	-2	-1	0	S^c
(4,3)	-1	0	0	S^c
(4,4)	0	0	0	S
(4,6)	-2	0	-1	S^c
(5,1)	-1	0	1	S^c
(5,3)	0	1	1	S^c
(5,5)	0	0	0	S
(5,6)	-1	1	0	S^c
(6,2)	1	0	1	S
(6,4)	2	0	1	S
(6,5)	1	-1	0	S
(6,6)	0	0	0	S

The I -lower approximations, the I -upper approximations, and the I -boundaries of S and S^c obtained by means of multi-graded dominance relations are as follows:

- $\underline{I}(S) = \{(1, 2), (1, 4), (1, 5), (3, 4), (6, 2), (6, 4), (6, 5)\}$.
- $\bar{I}(S) = \{(1, 1), (1, 2), (1, 4), (1, 5), (2, 2), (2, 3), (3, 2), (3, 3), (3, 4), (3, 5), (4, 4), (5, 3), (5, 5), (6, 2), (6, 4), (6, 5), (6, 6)\}$.
- $\underline{I}(S^c) = \{(2, 1), (2, 6), (4, 1), (4, 3), (4, 6), (5, 1), (5, 6)\}$.
- $\bar{I}(S^c) = \{(1, 1), (2, 1), (2, 2), (2, 3), (2, 6), (3, 2), (3, 3), (3, 5), (4, 1), (4, 3), (4, 4), (4, 6), (5, 1), (5, 3), (5, 5), (5, 6), (6, 6)\}$.
- $Bn_I(S) = Bn_I(S^c) = \{(1, 1), (2, 2), (2, 3), (3, 2), (3, 3), (3, 5), (4, 4), (5, 3), (5, 5), (6, 6)\}$.

Therefore, the quality of approximation is equal to 0.58. Moreover, there is only one reduct which is also the core, i.e., $RED_S(I) = CORE_S(I) = \{1\}$.

Finally, the following decision rules can be induced (within parentheses there are the pairs of objects supporting the rule):

- if $xP_1^{\geq 1}y$, then xSy (or, in words, if x is at least weakly preferred to y with respect to f_1 , then x outranks y),
(1,2),(1,4),(1,5),(3,4),(6,2),(6,4), (6,5)).
- if $xP_1^{\leq -1}y$, then $xScy$ (or, in words, if y is at least weakly preferred to x with respect to f_1 , then x does not outrank y),
(2,1),(2,6),(4,1),(4,3),(4,6), (5,1),(5,6)).
- if $xP_1^{\geq 0}y$, and $xP_1^{\leq 1}y$, (i.e., if xP_1^0y), then xSy or $xScy$ (or, in words, if x and y are indifferent with respect to f_1 , then x outranks y or x does not outrank y),
(1,1),(2,2),(2,3),(3,2),(3,3),(3,5),(4,4),
(5,3),(5,5),(6,6)).

Let us assume now that the DM accepts to express preferences with respect to criteria f_1, f_2, f_3 on an *ordinal scale* of preference, for which there is only information about a partial preorder on all possible pairs of evaluations. In this case, S and S^c can be approximated in the way described in subsection “Decision Rules,” i.e., without considering degrees of preference. The I -lower approximations, the I -upper approximations, and the I -boundaries of S and S^c are as follows:

- $\underline{I}(S) = \{(1,1), (1,2), (1,4), (1,5), (3,4), (4,4), (6,2), (6,4), (6,5), (6,6)\}$.
- $\overline{I}(S) = \{(1,1), (1,2), (1,4), (1,5), (2,2), (2,3), (3,2), (3,3), (3,4), (3,5), (4,4), (5,3), (5,5), (6,2), (6,4), (6,5), (6,6)\}$.
- $\underline{I}(S^c) = \{(2,1), (2,6), (4,1), (4,3), (4,6), (5,1), (5,6)\}$.
- $\overline{I}(S^c) = \{(2,1), (2,2), (2,3), (2,6), (3,2), (3,3), (3,5), (4,1), (4,3), (4,6), (5,1), (5,3), (5,5), (5,6)\}$.
- $Bn_I(S) = Bn_I(S^c) = \{(2,2), (2,3), (3,2), (3,3), (3,5), (5,3), (5,5)\}$.

Let us observe that the pairs (1,1), (4,4), and (6,6) belong now to the I -lower approximation of S and are not contained in the I -boundaries. Therefore, the quality of approximation is equal to 0.71. Moreover, there is still only one reduct which is also the core, i.e., again $RED_S(I) = CORE_S(I) = \{1\}$.

The following decision rules are induced from the above approximations and boundaries (within parentheses there are the pairs of objects supporting the rule):

- if $f_1(x)$ is at least high and $f_1(y)$ is at most high, then xSy , ((1,1),(1,2),(1,4),(1,5),(6,2),(6,4), (6,5),(6,6)).
- if $f_1(x)$ is at least low and $f_1(y)$ is at most low, then xSy , ((1,4),(3,4),(4,4),(6,4)).
- if $f_1(x)$ is at most medium and $f_1(y)$ is at least high, then $xScy$, ((2,1),(2,6),(4,1),(4,6),(5,1), (5,6)).
- if $f_1(x)$ is at most low and $f_1(y)$ is at least medium, then $xScy$, ((4,1),(4,3),(4,6)).
- if $f_1(x)$ is at least medium and $f_1(y)$ is at most medium and $f_1(x)$ is at most medium and $f_1(y)$ is at least medium, (i.e., if $f_1(x)$ is equal to medium and $f_1(y)$ is equal to medium), then xSy or $xScy$, ((2,2),(2,3),(3,2),(3,3),(3,5),(5,3), (5,5)).

DRSA for Decision Under Uncertainty

Basic Concepts

To apply the rough set approach to decision under uncertainty, the following basic elements must be considered:

- A set $S = \{s_1, s_2, \dots, s_u\}$ of states of the world, or simply *states*, which are supposed to be mutually exclusive and collectively exhaustive.
- An a priori probability distribution P over the states of the world: more precisely, the probabilities of states $s_1, s_2, \dots, s_u$ are $p_1, p_2, \dots, p_u$, respectively ($p_1 + p_2 + \dots + p_u = 1, p_i \geq 0, i = 1, \dots, u$).
- A set $A = \{a_1, a_2, \dots, a_m\}$ of *acts*.

- A set $X = \{x_1, x_2, \dots, x_r\}$ of *outcomes* or *consequences* expressed in monetary terms ($X \subseteq \mathfrak{R}$).
- A function $g: A \rightarrow X$ assigning to each pair act-state $(a_i, s_j) \in A \times S$ an outcome $x \in X$.
- A set of *classes* $\mathbf{CI} = \{Cl_1, Cl_2, \dots, Cl_p\}$, such that $Cl_1 \cup Cl_2 \cup \dots \cup Cl_p = A$, $Cl_r \cap Cl_q = \emptyset$ for each $r, q \in \{1, \dots, p\}$ with $r \neq q$; the classes of $\mathbf{CI}$ are preference-ordered according to the increasing order of their indices.
- A function $e: \rightarrow \mathbf{CI}$ assigning each act $a_i \in A$ to a class $Cl_j \in \mathbf{CI}$.

In this context, two different types of dominance can be considered:

1. *Classical dominance*: given $a_p, a_q \in A$, a_p dominates a_q iff, for each possible state of the world, act a_p gives an outcome at least as good as act a_q ; more formally, $g(a_p, s_j) \geq g(a_q, s_j)$, for each $s_j \in S$.
2. *Stochastic dominance*: given $a_p, a_q \in A$, a_p dominates a_q iff, for each outcome $x \in X$, act a_p gives an outcome at least as good as x with a probability at least as great as the probability that act a_q gives the same outcome, i.e., for all $x \in X$,

$$P[S(a_p, x)] \geq P[S(a_q, x)]$$

where, for each $(a_i, x) \in A \times X$, $S(a_i, x) = \{s_j \in S: g(a_i, s_j) \geq x\}$.

In Greco et al. (2001b), it has been shown how to apply stochastic dominance in this context. On the basis of an a priori probability distribution P , one can assign to each subset of states of the world $W \subseteq S (W \neq \emptyset)$ the probability $P(W)$ that one of the states in W is verified, i.e., $P(W) = \sum_{i: s_j \in W} p_{ij}$, and then to build up the set Π of all possible values $P(W)$, i.e.,

$$\Pi = \{\pi \in [0, 1] : \pi = P(W), W \subseteq S\}.$$

Let us define the following functions $z: A \times S \rightarrow \Pi$ and $z': A \times S \rightarrow \Pi$ assigning to each act-state pair $(a_i, s_j) \in A \times S$ a probability $\pi \in \Pi$, as follows:

$$z(a_i, s_j) = \sum_{r: g(a_i, s_r) \geq g(a_i, s_j)} p_r,$$

and

$$z'(a_i, s_j) = \sum_{r: g(a_i, s_r) \leq g(a_i, s_j)} p_r.$$

Therefore, $z(a_i, s_j)$ represents the probability of obtaining an outcome whose value is at least $g(a_i, s_j)$ by act a_i . Analogously, $z'(a_i, s_j)$ represents the probability of obtaining an outcome whose value is at most $g(a_i, s_j)$ by act a_i . On the basis of function $z(a_i, s_j)$, function $\rho: A \times \Pi \rightarrow X$ can be defined as follows:

$$\rho(a_i, \pi) = \max_{j: z(a_i, s_j) \geq \pi} g(a_i, s_j).$$

Thus, $\rho(a_i, \pi) = x$ means that the outcome got by act a_i is greater than or equal to x with a probability at least π (i.e., a probability π or greater). On the basis of function $z'(a_i, s_j)$, the function $\rho': A \times \Pi \rightarrow X$ can be defined as follows:

$$\rho'(a_i, \pi) = \min_{j: z'(a_i, s_j) \geq \pi} g(a_i, s_j).$$

$\rho'(a_i, \pi) = x$ means that the outcome got by act a_i is smaller than or equal to x with a probability at least π .

Let us observe that information given by $\rho(a_i, \pi)$ and $\rho'(a_i, \pi)$ is related. In fact, if the elements of Π , $0 = \pi_{(0)}, \pi_{(1)}, \pi_{(2)}, \dots, \pi_{(d)} = 1$ ($d = |\Pi|$), are reordered in such a way that $0 = \pi_{(0)} \leq \pi_{(1)} \leq \pi_{(2)} \leq \dots \leq \pi_{(d)} = 1$, then

$$\rho(a_i, \pi_{(j)}) = \rho'(a_i, 1 - \pi_{(j-1)}).$$

Therefore, $\rho(a_i, \pi_{(j)}) \leq x$ is equivalent to $\rho'(a_i, 1 - \pi_{(j-1)}) \geq x$, $a_i \in A$, $\pi_{(j-1)}, \pi_{(j)} \in \Pi$, $x \in X$. This implies that the analysis of the possible decisions can be equivalently conducted on values of either $\rho(a_i, \pi)$ or $\rho'(a_i, \pi)$. However, from the point of view of representation of results, it is interesting to consider both values $\rho(a_i, \pi)$ and $\rho'(a_i, \pi)$. The reason is that, contrary to intuition,

$\rho(a_i, \pi) \leq x$ is not equivalent to the statement that by act a_i the outcome is smaller than or equal to x with a probability at least π . The following example clarifies this point. Let us consider game a with rolling a dice, in which if the result is 1, then the gain is \$1, if the result is 2 then the gain is \$2, and so on. Suppose, moreover, that the dice is equilibrated and thus each result is equiprobable with probability 1/6. The values of $\rho(a_i, \pi)$ for all possible values of probability are:

$$\begin{aligned} \rho(a, 1/6) &= \$6, & \rho(a, 2/6) &= \$5, \\ \rho(a, 3/6) &= \$4, & \rho(a, 4/6) &= \$3, \\ \rho(a, 5/6) &= \$2, & \rho(a, 6/6) &= \$1. \end{aligned}$$

Let us remark that $\rho(a, 4/6) = \$3$ is not equivalent to $\rho'(a, 2/6) = \$3$. In fact,

$$\begin{aligned} \rho'(a, 1/6) &= \$1, & \rho'(a, 2/6) &= \$2, \\ \rho'(a, 3/6) &= \$3, & \rho'(a, 4/6) &= \$4, \\ \rho'(a, 5/6) &= \$5, & \rho'(a, 6/6) &= \$6, \end{aligned}$$

and therefore $\rho(a, 4/6) = \$3$ implies $\rho'(a, 3/6) = \$3$.

In the context of stochastic acts, an outcome expressed in positive terms refers to $\rho(a, \pi)$ giving a lower bound of an outcome ("for act a there is a probability π to gain at least $\rho(a, \pi)$ "), while an outcome expressed in negative terms refers to $\rho'(a, \pi)$ giving an upper bound of an outcome ("for act a there is a probability π to gain at most $\rho'(a, \pi)$ ").

Given $a_p, a_q \in A$, a_p stochastically dominates a_q if and only if $\rho(a_p, \pi) \geq \rho(a_q, \pi)$ for each $\pi \in \Pi$. This is equivalent to the statement: given $a_p, a_q \in A$, a_p stochastically dominates a_q if and only if $\rho'(a_p, \pi) \leq \rho'(a_q, \pi)$ for each $\pi \in \Pi$.

For example, consider the game a^* with rolling a dice, in which if the result is 1, then the gain is \$7, if the result is 2 then the gain is \$6, and so on until the case in which the result is 6 and the gain is \$2. In this case, game a^* stochastically dominates game a because $\rho(a^*, 1/6) = \$7$ is not smaller than $\rho(a, 1/6) = \$6$, $\rho(a^*, 2/6) = \$6$ is not smaller than $\rho(a, 2/6) = \$5$, and so on. Equivalently, game a^* stochastically dominates game a because $\rho'(a^*, 1/6) = \$2$ is not

smaller than $\rho'(a, 1/6) = \$1$, $\rho'(a^*, 2/6) = \$3$ is not smaller than $\rho'(a, 2/6) = \$2$, and so on.

DRSA can be applied in the context of decision under uncertainty considering as set of objects U the set of acts A , as set of criteria (condition attributes) I the set Π , as decision attribute $\{d\}$ the classification CI , as value set of all criteria the set X , as information function f a function f such that $f(a_i, \pi) = \rho(a_i, \pi)$ and $f(a_i, cl) = e(a_i)$. Let us observe that due to equivalence $\rho(a_i, \pi_{(j)}) = \rho'(a_i, 1 - \pi_{(j-1)})$, one can also consider information function $f'(a_i, \pi) = \rho'(a_i, \pi)$.

The aim of the rough set approach to preferences under uncertainty is to explain the preferences of the DM represented by the assignments of the acts from A to the classes from CI in terms of stochastic dominance, expressed by means of function ρ . The syntax of decision rules obtained from this rough set approach is as follows:

1. $D_{\geq}$ -decision rules with the following syntax:
 "if $\rho(a, p_{h_1}) \geq x_{h_1}$ and ... , and $\rho(a, p_{h_z}) \geq x_{h_z}$, then $a \in Cl_r^{\geq}$ (i.e., "if by act a the outcome is at least x_{h_1} with probability at least p_{h_1} , and ... , and the outcome is at least x_{h_z} with probability at least p_{h_z} , then $a \in Cl_r^{\geq}$ "), where $p_{h_1}, \dots, p_{h_z} \in \Pi$, $x_{h_1}, \dots, x_{h_z} \in X$, and $r \in \{2, \dots, p\}$.
2. $D_{\leq}$ -decision rules with the following syntax:
 "if $\rho'(a, p_{h_1}) \leq x_{h_1}$ and ... , and $\rho'(a, p_{h_z}) \leq x_{h_z}$, then $a \in Cl_r^{\leq}$ (i.e., "if by act a the outcome is at most x_{h_1} with probability at least p_{h_1} , and ... , and the outcome is at most x_{h_z} with probability at least p_{h_z} , then $a \in Cl_r^{\leq}$ "), where $p_{h_1}, \dots, p_{h_z} \in \Pi$, $x_{h_1}, \dots, x_{h_z} \in X$ and $r \in \{1, \dots, p-1\}$.
3. $D_{\geq \leq}$ -decision rules with the following syntax:
 "if $\rho(a, p_{h_1}) \geq x_{h_1}$ and ... , and $\rho(a, p_{h_z}) \geq x_{h_z}$ and $\rho'(a, p_{h_{w+1}}) \leq x_{h_{w+1}}$ and ... , and $\rho'(a, p_{h_t}) \leq x_{h_t}$, then $a \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$ (i.e., "if by act a the outcome is at least x_{h_1} with probability at least p_{h_1} , and ... , and the outcome is at least x_{h_w} with probability at least p_{h_w} and the outcome is at most $x_{h_{w+1}}$ with probability at least $p_{h_{w+1}}$, and ... , and the

outcome is at most x_{h_z} with probability at least p_{h_z} , then $a \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$ ", where $p_{h_1}, \dots, p_{h_w}, p_{h_{w+1}}, \dots, p_{h_z} \in \Pi, x_{h_1}, \dots, x_{h_z} \in X$ and $s, t \in \{1, \dots, p\}$, such that $s < t$.

According to the meaning of $\rho(a_i, p)$ and $\rho'(a_i, p)$ discussed above, $D_{\geq}$ -decision rules are expressed in terms of $\rho(a_i, p)$, $D_{\leq}$ -decision rules are expressed in terms of $\rho'(a_i, p)$, and $D_{\geq \leq}$ -decision rules are expressed in terms of both $\rho(a_i, p)$ and $\rho'(a_i, p)$. Let us observe that due to equivalence $\rho(a_i, \pi_{(j)}) = \rho'(a_i, 1 - \pi_{(j-1)})$, all above decision rules can be expressed equivalently in terms of values of $\rho(a_i, p)$ or $\rho'(a_i, p)$. For example, a $D_{\geq}$ -decision rule $r_{\geq}(\rho) =$ "if $\rho(a, p_{h_1}) \geq x_{h_1}$ and $\dots$, and $\rho(a, p_{h_z}) \geq x_{h_z}$, then $a \in Cl_r^{\geq}$ " can be expressed in terms of $\rho'(a_i, p)$ as $r_{\geq}(\rho') =$ "if $\rho'(a, p_{h_1}^*) \geq x_{h_1}$ and $\dots$, and $\rho'(a, p_{h_z}^*) \geq x_{h_z}$, then $a \in Cl_r^{\geq}$ " where, if $p_{h_r} = \pi_{j_r}$, then $p_{h_r}^* = 1 - \pi_{j_r-1}$, with $r = 1, \dots, z$, and $0 = \pi(0), \pi(1), \pi(2), \dots, \pi(|P|) = 1$ reordered in such a way that $0 = \pi(0) \leq \pi(1) \leq \pi(2) \leq \dots \leq \pi(|P|) = 1$. Analogously, a $D_{\leq}$ -decision rule $r_{\leq}(\rho') =$ "if $\rho'(a, p_{h_1}) \leq x_{h_1}$ and $\dots$, and $\rho'(a, p_{h_z}) \leq x_{h_z}$, then $a \in Cl_r^{\leq}$ " can be expressed in terms of $\rho(a_i, p)$ as $r_{\leq}(\rho) =$ "if $\rho(a, p_{h_1}^*) \leq x_{h_1}$ and $\dots$, and $\rho(a, p_{h_z}^*) \leq x_{h_z}$, then $a \in Cl_r^{\leq}$ " where, if $p_{h_r} = \pi_{j_r}$, then $p_{h_r}^* = 1 - \pi_{j_r-1}'$, with $r = 1, \dots, z$, and $0 = \pi(0), \pi(1), \pi(2), \dots, \pi(|P|) = 1$ reordered in such a way that $0 = \pi(0) \leq \pi(1) \leq \pi(2) \leq \dots \leq \pi(|P|) = 1$.

Let us observe, however, that $r_{\geq}(\rho)$ is an expression much more natural and meaningful than $r_{\geq}(\rho')$, as well as $r_{\leq}(\rho')$ is an expression much more natural and meaningful than $r_{\leq}(\rho)$. Another useful remark concerns minimality of rules, related to the specific intrinsic structure of the stochastic dominance. Let us consider the following two decision rules:

- $r1 \equiv$ "if by act a the outcome is at least 100 with probability at least 0.25, then a is at least good."
- $r2 \equiv$ "if by act a the outcome is at least 100 with probability at least 0.50, then a is at least good."

$r1$ and $r2$ can be induced from the analysis of the same information table, because they involve different criteria (condition attributes). In fact, $r1$ involves attribute $\rho(a, 0.25)$ (it can be expressed as "if $\rho(a, 0.25) \geq 100$, then a is at least good"), $r2$ involves attribute $\rho(a, 0.50)$ (it can be expressed as "if $\rho(a, 0.50) \geq 100$, then a is at least good"). Considering the structure of the stochastic dominance, the condition part of rule $r1$ is the weakest. In fact, rule $r1$ requires a cumulated outcome to be at least 100 with probability of 0.25, while rule $r2$ requires the same outcome but with a greater probability, 0.5 against 0.25. Since the decision part of these two rules is the same, $r1$ is minimal among these two rules. From a practical point of view, this observation says that, if one induces decision rules using the algorithms designed for DRSA, it is necessary to further filter the obtained results in order to remove rules which are not minimal in the specific context of the DRSA analysis based on stochastic dominance.

To complete presentation of the methodological basis of DRSA applied to decisions under uncertainty, it is worth mentioning two important extensions. One concerns consideration of decision under uncertainty with outcomes distributed over time. In Greco et al. (2010a), we proposed a rough set model based on a combination of time dominance and stochastic dominance. The model is able to handle nonadditive probability distributions, and even qualitative ordinal distributions. DRSA gives also in this case a comprehensible representation of DM's time-dependent preferences under uncertainty in terms of decision rules. The second extension concerns application of DRSA adapted to decision under uncertainty within the Robust Ordinal Regression (ROR) methodology, presented in subsection "Robust Ordinal Regression for DRSA" for the case of ordinal classification. As explained in Kadziński et al. (2016), ROR-DRSA permits to take into account the whole family of minimal sets of decision rules induced from DM's preference information provided in terms of decision examples concerning classification of some acts characterized by probabilistic gains. In result, one obtains necessary and possible assignments of all considered acts to quality classes.

Example Illustrating DRSA in the Context of Decision Under Uncertainty

The following example illustrates the approach. Let us consider

- A set $S = \{s_1, s_2, s_3\}$ of states of the world.
- An a priori probability distribution P over the states of the world defined as follows: $p_1 = 0.25$, $p_2 = 0.35$, $p_3 = 0.40$.
- A set $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ of acts.
- A set $X = \{0, 10, 15, 20, 30\}$ of consequences.
- A set of classes $CI = \{Cl_1, Cl_2, Cl_3\}$, where Cl_1 is the set of bad acts, Cl_2 is the set of medium acts, and Cl_3 is the set of good acts.
- A function $g: A \rightarrow X$ assigning to each act-state pair $(a_i, s_j) \in A \times S$ a consequence $x_h \in X$ and a function $e: A \rightarrow CI$ assigning each act $a_i \in A$ to a class $Cl_j \in CI$ presented in the following Table 9.

Table 10 shows the values of function $\rho(a_i, p)$. Table 10 is the decision table on which the DRSA is applied. Let us give some examples of the interpretation of the values in Table 10. The column of act a_3 can be read as follow:

- The value 20 in the row corresponding to 0.25 means that the outcome is at least 20 with a probability of at least 0.25.
- The value 15 in the row corresponding to 0.60 means that the outcome is at least 15 with a probability of at least 0.60.
- The value 0 in the row corresponding to 0.75 means that the outcome is at least 0 with a probability of at least 0.75.

Analogously, the row corresponding to 0.65 can be read as follows:

- The value 10 relative to a_1 , means that by act a_1 , the outcome is at least 10 with a probability of at least 0.65.
- The value 20 relative to a_2 , means that by act a_2 , the outcome is at least 20 with a probability of at least 0.65.
- And so on.

Applying DRSA, the following upward union and downward union of classes are approximated:

- $Cl_2^{\geq} = Cl_2 \cup Cl_3$, i.e., the set of the acts at least medium

Rough Sets in Decision-Making, Table 9 Acts, consequences, and assignment to classes from CI

	p_j	a_1	a_2	a_3	a_4	a_5	a_6
s_1	0.25	30	0	15	0	20	10
s_2	0.35	10	20	0	15	10	20
s_3	0.40	10	20	20	20	20	20
CI		Good	Medium	Medium	Bad	Medium	Good

Rough Sets in Decision-Making, Table 10 Acts, values of function $\rho(a_i, p)$, and assignment to classes from CI

	a_1	a_2	a_3	a_4	a_5	a_6
0.25	30	20	20	20	20	20
0.35	10	20	20	20	20	20
0.40	10	20	20	20	20	20
0.60	10	20	15	15	20	20
0.65	10	20	20	20	20	20
0.75	10	20	0	15	10	20
1	10	0	0	0	10	10
CI	Good	Medium	Medium	Bad	Medium	Good

- $Cl_3^{\geq} = Cl_3$, i.e., the set of the acts (at least) good
- $Cl_1^{\leq} = Cl_1$, i.e., the set of the acts (at most) bad
- $Cl_2^{\leq} = Cl_1 \cap Cl_2$, i.e., the set of the acts at most medium

The first result of the DRSA approach was a discovery that the decision table (Table 10) is not consistent. Indeed, Table 10 shows that act a_4 stochastically dominates act a_3 ; however, act a_3 is assigned to a better class (medium) than act a_4 (bad). Therefore, act a_3 cannot be assigned without doubts to the set of the class of the at least medium acts as well as act a_4 cannot be assigned without doubts to the set of the classes of the (at most) bad acts. In consequence, lower approximation and upper approximation of $Cl_2^{\geq}$, $Cl_3^{\geq}$ and $Cl_1^{\leq}$, $Cl_2^{\leq}$ are equal, respectively, to

- $\underline{I}(Cl_2^{\geq}) = \{a_1, a_2, a_5, a_6\} = Cl_2^{\geq} - \{a_3\}$.
- $\bar{I}(Cl_2^{\geq}) = \{a_1, a_2, a_5, a_6\} = Cl_2^{\geq} \cup \{a_4\}$.
- $\underline{I}(Cl_3^{\geq}) = \{a_1, a_6\} = Cl_3^{\geq}$.
- $\bar{I}(Cl_3^{\geq}) = \{a_1, a_6\} = Cl_3^{\geq}$.
- $\underline{I}(Cl_1^{\leq}) = \emptyset = Cl_1^{\leq} - \{a_4\}$.
- $\bar{I}(Cl_1^{\leq}) = \{a_3, a_4\} = Cl_1^{\leq} \cup \{a_3\}$.
- $\underline{I}(Cl_2^{\leq}) = \{a_2, a_3, a_4, a_5\} = Cl_2^{\leq}$.
- $\bar{I}(Cl_2^{\leq}) = \{a_2, a_3, a_4, a_5\} = Cl_2^{\leq}$.

Since there are two inconsistent acts (a_3, a_4) on a total of six acts, then the quality of approximation of the ordinal classification is equal to 4/6. The second discovery was one reduct of criteria (condition attributes) ensuring the same quality of approximation as the whole set Π of probabilities: $RED_{CI} = \{0.25, 0.75, 1\}$. This means that the preferences of the DM can be explained using only the probabilities from RED_{CI} . RED_{CI} is also the core because no probability value in RED_{CI} can be removed without deteriorating the quality of approximation. The third discovery was a set of minimal decision rules describing the DM's preferences (within parentheses there is a verbal interpretation of the corresponding decision rule, and the supporting acts):

1. if $\rho(a_i, 0.25) \geq 30$, then $a_i \in Cl_3^{\geq}$
(if the probability of gaining at least 30 is at least 0.25, then act a_i is at least good) (a_1).
2. if $\rho(a_i, 0.75) \geq 20$ and $\rho(a_i, 1) \geq 10$, then $a_i \in Cl_3^{\geq}$
(if the probability of gaining at least 20 is at least 0.75 and the probability of gaining at least 10 is (at least) 1 (i.e., for sure the gaining is at least 10), then act a_i is at least good) (a_6).
3. if $\rho(a_i, 1) \geq 10$, then $a_i \in Cl_2^{\geq}$
(if the probability of gaining at least 10 is (at least) 1 (i.e., for sure the gaining is at least 10), then act a_i is at least medium) (a_1, a_5, a_6).
4. if $\rho(a_i, 0.75) \geq 20$, then $a_i \in Cl_2^{\geq}$
(if the probability of gaining at least 20 is at least 0.75, then act a_i is at least medium) (a_2, a_6).
5. if $\rho(a_i, 0.25) \leq 20$ (i.e., $\rho'(a_i, 1) \leq 20$) and $\rho(a_i, 0.75) \leq 15$ (i.e., $\rho'(a_i, 0.35) \leq 15$), then $a_i \in Cl_2^{\leq}$
(if the probability of gaining at most 20 is (at least) 1 (i.e., for sure you gain at most 20) and the probability to gain at most 15 is at least 0.35, then act a_i is at most medium) (a_3, a_4, a_5).
6. if $\rho(a_i, 1) \leq 0$ (i.e., $\rho'(a_i, 0.25) \leq 0$), then $a_i \in Cl_2^{\leq}$
(if the probability of gaining at most 0 is at least 0.25, then act a_i is at most medium) (a_2, a_3, a_4).
7. if $\rho(a_i, 1) \geq 0$ and $\rho'(a_i, 1) \leq 0$ (i.e., $\rho(a_i, 1) = 0$) and $\rho(a_i, 0.75) \geq 15$ ($\rho'(a_i, 0.35) \leq 15$), then $a_i \in Cl_1 \cup Cl_2$
(if the probability of gaining at least 0 is 1, then act a_i is at most medium) (a_2, a_3, a_4).

Minimal sets of minimal decision rules represent the most concise and nonredundant knowledge contained in Table 9 (and, consequently, in Table 10). The above minimal set of 7 decision rules uses 3 attributes (probability 0.25, 0.75 and 1) and 11 elementary conditions, i.e., 26% of descriptors from the original data table (Table 10). Of course, this is only a didactic example: representation in terms of decision rules of larger sets of exemplary acts from real applications are more synthetic in the sense of the percentage of used descriptors from the original decision table.

Multiple Criteria Decision Analysis Using Association Rules

In multiple criteria decision analysis, the DM is often interested in relationships between attainable values of criteria. This information is particularly useful in multiobjective optimization (see Greco et al. 2008a and section “Interactive Multiobjective Optimization Using DRSA (IMO-DRSA)”). For instance, in a car selection problem, one can observe that in the set of considered cars, if the maximum speed is at least 200 km/h and the time to reach 100 km/h is at most 7 s, then the price is not less than 40,000\$ and the fuel consumption is not less than 9 l per 100 km. These relationships are association rules whose general syntax, in case of minimization of criteria f_i , $i \in I$, is:

“if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then $f_{i_{p+1}}(x) \geq r_{i_{p+1}}$ and ... and $f_{i_q}(x) \geq r_{i_q}$ ”, where $\{i_1, \dots, i_q\} \subseteq I$, $r_{i_1}, \dots, r_{i_q} \in \mathfrak{R}$.

If criterion f_i , $i \in I$, should be maximized, the corresponding condition in the association rule should be reversed, i.e., in the premise, the condition becomes $f_i(x) \geq r_i$, and in the conclusion it becomes $f_i(x) \leq r_i$.

Given an association rule $r \equiv$ “if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then $f_{i_{p+1}}(x) \geq r_{i_{p+1}}$ and ... and $f_{i_q}(x) \geq r_{i_q}$ ”, an object $y \in U$ supports r if $f_{i_1}(y) \leq r_{i_1}$ and ... and $f_{i_p}(y) \leq r_{i_p}$ and $f_{i_{p+1}}(y) \geq r_{i_{p+1}}$ and ... and $f_{i_q}(y) \geq r_{i_q}$. Moreover, object $y \in U$ supporting decision rule r is a *base* of r if $f_{i_1}(y) = r_{i_1}$ and ... and $f_{i_p}(y) = r_{i_p}$ and $f_{i_{p+1}}(y) = r_{i_{p+1}}$ and ... and $f_{i_q}(y) = r_{i_q}$. An association rule having at least one base is called *robust*.

An association rule $r \equiv$ “if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then $f_{i_{p+1}}(x) \geq r_{i_{p+1}}$ and ... and $f_{i_q}(x) \geq r_{i_q}$ ” holds in universe U if:

1. There is at least one $y \in U$ supporting r .
2. r is not contradicted in U , i.e., there is no $z \in U$ such that $f_{i_1}(z) \leq r_{i_1}$ and ... and $f_{i_p}(z) \leq r_{i_p}$, while *not* $f_{i_{p+1}}(z) \geq r_{i_{p+1}}$ or ... or $f_{i_q}(z) \geq r_{i_q}$.

Given the two association rules:

- $r_1 \equiv$ “if $f_{i_1}(x) \leq r_{i_1}^1$ and ... and $f_{i_p}(x) \leq r_{i_p}^1$, then $f_{i_{p+1}}(x) \geq r_{i_{p+1}}^1$ and ... and $f_{i_q}(x) \geq r_{i_q}^1$.”
- $r_2 \equiv$ “if $f_{j_1}(x) \leq r_{j_1}^2$ and ... and $f_{j_s}(x) \leq r_{j_s}^2$, then $f_{j_{s+1}}(x) \geq r_{j_{s+1}}^2$ and ... and $f_{j_t}(x) \geq r_{j_t}^2$.”

rule r_1 is not weaker than rule r_2 , denoted by $r_1 \succeq r_2$, if:

- $\alpha)$ $\{i_1, \dots, i_p\} \subseteq \{j_1, \dots, j_s\}$.
- $\beta)$ $r_{i_1}^1 \geq r_{i_1}^2, \dots, r_{i_p}^1 \geq r_{i_p}^2$.
- $\gamma)$ $\{i_{p+1}, \dots, i_q\} \supseteq \{j_{s+1}, \dots, j_t\}$.
- $\delta)$ $r_{j_{s+1}}^1 \geq r_{j_{s+1}}^2, \dots, r_{j_t}^1 \geq r_{j_t}^2$.

Conditions (β) and (δ) are formulated for criteria f_i to be minimized. If criterion f_i should be maximized, the corresponding inequalities should be reversed, i.e., $r_i^1 \leq r_i^2$ in condition (β) as well as in condition (δ) . Notice that $\succeq$ is a binary relation on the set of association rules, which is a partial preorder, i.e., it is reflexive (each rule is not weaker than itself) and transitive. The asymmetric part of the relation $\succeq$ is denoted by $\triangleright$, and $r_1 \triangleright r_2$ reads “ r_1 is stronger than r_2 ”.

For example, consider the following association rules:

- $r_1 \equiv$ “if the maximum speed is at least 200 km/h and the time to reach 100 km/h is at most 7 seconds, then the price is not less than 40,000\$ and the fuel consumption is not less than 9 liters per 100 km.”
- $r_2 \equiv$ “if the maximum speed is at least 200 km/h and the time to reach 100 km/h is at most 7 seconds and the horse power is at least 175 kW, then the price is not less than 40,000\$ and the fuel consumption is not less than 9 liters per 100 km.”
- $r_3 \equiv$ “if the maximum speed is at least 220 km/h and the time to reach 100 km/h is at most 7 seconds, then the price is not less than 40,000\$ and the fuel consumption is not less than 9 liters per 100 km.”
- $r_4 \equiv$ “if the maximum speed is at least 200 km/h and the time to reach 100 km/h is at

most 7 seconds, then the price is not less than 40,000\$.”

- $r_5 \equiv$ “if the maximum speed is at least 200 km/h and the time to reach 100 km/h is at most 7 seconds, then the price is not less than 35,000\$ and the fuel consumption is not less than 9 liters per 100 km.”
- $r_6 \equiv$ “if the maximum speed is at least 220 km/h and the time to reach 100 km/h is at most 7 seconds and the horse power is at least 175 kW, then the price is not less than 35,000\$.”

Let us observe that rule r_1 is stronger than each of the other five rules for the following reasons:

- $r_1 \triangleright r_2$ for condition (α) because, all things equal elsewhere, in the premise of r_2 there is an additional condition: “the horse power is at least 175 kW.”
- $r_1 \triangleright r_3$ for condition (β) because, all things equal elsewhere, in the premise of r_3 there is a condition with a worse threshold value: “the maximum speed is at least 220 km/h” instead of “the maximum speed is at least 200 km/h.”
- $r_1 \triangleright r_4$ for condition (γ) because, all things equal elsewhere, in the conclusion of r_4 one condition is missing: “the fuel consumption is not less than 9 liters per 100 km.”
- $r_1 \triangleright r_5$ for condition (δ) because, all things equal elsewhere, in the conclusion of r_5 there is a condition with a worse threshold value: “the price is not less than 35,000 \$” instead of “the price is not less than 40,000\$.”
- $r_1 \triangleright r_6$ for conditions (α), (β), (γ), and (δ) because all weak points for which rules r_2 , r_3 , r_4 , and r_5 are weaker than rule r_1 are present in r_6 .

An association rule r is *minimal* if there is no other rule stronger than r with respect to $\triangleright$. An algorithm for induction of association rules from preference ordered data has been presented in Greco et al. (2002c).

Interactive Multiobjective Optimization Using DRSA (IMO-DRSA)

This section presents a recently proposed method for interactive multiobjective optimization using dominance-based rough set approach (IMO-DRSA) (Greco et al. 2008a). The method is composed of the following steps.

Step 1. Generate a representative sample of solutions from the currently considered part of the Pareto optimal set.

Step 2. Present the sample to the DM, possibly together with association rules showing relationships between attainable values of objective functions in the Pareto optimal set.

Step 3. If the DM is satisfied with one solution from the sample, then this is the most preferred solution and the procedure stops. Otherwise continue.

Step 4. Ask the DM to indicate a subset of relatively “good” solutions in the sample.

Step 5. Apply DRSA to the current sample of solutions classified into “good” and “others” solutions, in order to induce a set of decision rules with the following syntax “if $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}$ and ... and $f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$, then solution $\mathbf{x}$ is good,” $\{j_1, \dots, j_p\} \subseteq \{1, \dots, n\}$.

Step 6. Present the obtained set of rules to the DM.

Step 7. Ask the DM to select the decision rules most adequate to his/her preferences.

Step 8. Adjoin the constraints $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}, \dots, f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$ coming from the rules selected in Step 7 to the set of constraints imposed on the Pareto optimal set, in order to focus on a part interesting from the point of view of DM’s preferences.

Step 9. Go back to Step 1.

In a sequence of iterations, the method is exploring the Pareto optimal set of a multiobjective optimization problem or an approximation of this set. In the calculation stage (Step 1), any multiobjective optimization method, which finds the Pareto optimal set or its approximation, such as evolutionary multiobjective optimization methods, can be used. In the dialogue stage of the

method (Steps 2–7), the DM is asked to select a decision rule induced from his/her preference information, which is equivalent to fixing some upper bounds for the minimized objective functions f_j .

In Step 1, the representative sample of solutions from the currently considered part of the Pareto optimal set can be generated using one of existing procedures, such as (Jaszkiewicz and Słowiński 1999; Steuer and Choo 1983; Wierzbicki 1980). It is recommended to use a fine-grained sample of representative solutions to induce association rules; however, the sample of solutions presented to the DM in Step 2 should be much smaller (about a dozen) in order to avoid an excessive cognitive effort of the DM. Otherwise, the DM would risk to give non reliable information.

The association rules presented in Step 2 help the DM in understanding what (s)he can expect from the optimization problem. More precisely, any association rule

“if $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then $f_{i_{p+1}}(x) \geq r_{i_{p+1}}$ and ... and $f_{i_q}(x) \geq r_{i_q}$ ”,
where $\{i_1, \dots, i_q\} \subseteq I, r_{i_1}, \dots, r_{i_q} \in \mathfrak{R}$

says to the DM that, if (s)he wants attain the values of objective functions $f_{i_1}(x) \leq r_{i_1}$ and ... and $f_{i_p}(x) \leq r_{i_p}$, then (s)he cannot reasonably expect to obtain values of objective functions $f_{i_{p+1}}(x) < r_{i_{p+1}}$ and ... and $f_{i_q}(x) < r_{i_q}$.

With respect to the ordinal classification of solutions into the two classes of “good” and “others,” observe that “good” means in fact “relatively good,” i.e., better than the rest. In case, the DM would refuse to classify as “good” any solution, one can ask the DM to specify some minimal requirements of the type $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}$ and ... and $f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$ for “good” solutions. These minimal requirements give some constraints that can be used in Step 8, in the same way as the analogous constraints coming from selected decisions rules.

The rules considered in Step 5 have a syntax corresponding to minimization of objective functions. In case of maximization of an objective function f_j , the condition concerning this objective

in the decision rule should have the form $f_j(\mathbf{x}) \geq \alpha_j$.

Remark, moreover, that the Pareto optimal set reduced in Step 8 by constraints $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}, \dots, f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$ is certainly not empty if these constraints are coming from one decision rule only. Since robust rules (see subsection “Dominance-Based Rough Set Approach (DRSA)”) are considered, the threshold values $\alpha_{j_1}, \dots, \alpha_{j_p}$ are values of objective functions of some solutions from the Pareto optimal set. If $\{j_1, \dots, j_p\} = \{1, \dots, n\}$, i.e., $\{j_1, \dots, j_p\}$ is the set of all objective functions, then the new reduced part of the Pareto optimal set contains only one solution $\mathbf{x}$ such that $f_1(\mathbf{x}) = \alpha_1, \dots, f_n(\mathbf{x}) = \alpha_n$. If $\{j_1, \dots, j_p\} \subset \{1, \dots, n\}$, i.e., $\{j_1, \dots, j_p\}$ is a proper subset of the set of all objective functions, then the new reduced part of the Pareto optimal set contains solutions satisfying conditions $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}$ and ... and $f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$. Since the considered rules are robust, then there is at least one solution $\mathbf{x}$ satisfying these constraints. When the Pareto optimal set is reduced in Step 8 by constraints $f_{j_1}(\mathbf{x}) \leq \alpha_{j_1}, \dots, f_{j_p}(\mathbf{x}) \leq \alpha_{j_p}$ coming from more than one rule, then it is possible that the resulting reduced part of the Pareto optimal set is empty. Thus, before passing to Step 9, it is necessary to verify if the reduced Pareto optimal set is not empty. If the reduced Pareto optimal set is empty, then the DM is required to revise his/her selection of rules. The DM can be supported in this task, by information about minimal sets of constraints $f_j(\mathbf{x}) \leq \alpha_j$ coming from the considered decision rules to be removed in order to get a non-empty part of the Pareto optimal set.

The constraints introduced in Step 8 are maintained in the following iterations of the procedure; however, they cannot be considered as irreversible. Indeed, the DM can come back to the Pareto optimal set considered in one of previous iterations and continue from this point. This is in the spirit of a learning oriented conception of interactive multiobjective optimization, i.e., it agrees with the idea that the interactive procedure permits the DM to learn about his/her preferences and about the “shape” of the Pareto optimal set.

Example Illustrating IMO-DRSA in the Context of Multiobjective Optimization

To illustrate the interactive multiobjective optimization procedure based on DRSA, a product mix problem is considered. There are three products: A, B, C which are produced in quantities denoted by x_A, x_B , and x_C , respectively. The unit prices of the three products are $p_A = 20$, $p_B = 30$, and $p_C = 25$. The production process involves two machines. The production times of A, B, C on the first machine are equal to $t_{1A} = 5$, $t_{1B} = 8$, and $t_{1C} = 10$, and on the second machine, they are equal to $t_{2A} = 8$, $t_{2B} = 6$, and $t_{2C} = 2$. Two raw materials are used in the production process. The first raw material has a unit cost of 6 and the quantity required for production of one unit of A, B , and C is $r_{1A} = 1$, $r_{1B} = 2$, and $r_{1C} = 0.75$, respectively. The second raw material has a unit cost of 8 and the quantity required for production of one unit of A, B , and C is $r_{2A} = 0.5$, $r_{2B} = 1$, and $r_{2C} = 0.5$, respectively. Moreover, the market cannot absorb a production greater than 10, 20, and 10 units for A, B , and C , respectively. To decide how much of A, B , and C should be produced, the following objectives have to be taken into account:

- Profit (to be maximized)
- Time (total production time on two machines – to be minimized)
- Production of A (to be maximized)
- Production of B (to be maximized)
- Production of C (to be maximized)
- Sales (to be maximized)

The above product mix problem can be formulated as the following multiobjective optimization problem:

Maximize

$$20x_A + 30x_B + 25x_C - (1x_A + 2x_B + 0.75x_C)6 - (0.5x_A + 1x_B + 0.5x_C)8 \quad [\text{Profit}],$$

$$\text{Minimize } 5x_A + 8x_B + 10x_C + 8x_A + 6x_B + 2x_C \quad [\text{Time}],$$

$$\text{Maximize } x_A \quad [\text{Production of } A],$$

$$\text{Maximize } x_B \quad [\text{Production of } B],$$

$$\text{Maximize } x_C \quad [\text{Production of } C],$$

$$\text{Maximize } 20x_A + 30x_B + 25x_C \quad [\text{Sales}],$$

subject to:

$$x_A \leq 10, \quad x_B \leq 20, \quad x_C \leq 10$$

[Market absorption limits],

$$x_A \geq 0, \quad x_B \geq 0, \quad x_C \geq 0$$

[Non-negativity constraints].

A sample of representative Pareto optimal solutions has been calculated and proposed to the DM. Observe that the considered problem is a multiple objective linear programming (MOLP) problem, and thus representative Pareto optimal solutions can be calculated using classical linear programming looking for the solutions optimizing each one of the considered objectives or fixing all the considered objective functions but one at a satisfying value, and looking for the solution optimizing the remaining objective function. The set of representative Pareto optimal solutions is shown in Table 11. Moreover, a set of potentially interesting association rules have been induced from the sample and presented to the DM. These rules represent strongly supported relationships between attainable values of objective functions. The association rules are the following (between parentheses there are id's of solutions supporting the rule):

1. if Time ≤ 140 , then Profit ≤ 180.38 and Sales ≤ 280.77
(s1, s2, s3, s4, s12, s13).
2. if Time ≤ 150 , then Profit ≤ 188.08 and Sales ≤ 296.15
(s1, s2, s3, s4, s5, s6, s7, s8, s9, s12, s13).
3. if $x_B \geq 2$, then Profit ≤ 209.25 and $x_A \leq 6$ and $x_C \leq 7.83$
(s4, s5, s6, s9, s10, s11, s12, s13).
4. if Time ≤ 150 , then $x_B \leq 3$
(s1, s2, s3, s4, s5, s6, s7, s8, s9, s12, s13).
5. if Profit ≥ 148.38 and Time ≤ 150 , then $x_B \leq 2$
(s1, s2, s3, s5, s7, s8).
6. if $x_A \geq 5$, then Time ≥ 150
(s5, s6, s8, s9, s10, s11).

Rough Sets in Decision-Making, Table 11 A sample of Pareto optimal solutions proposed in the first iteration

Solution	Profit	Time	Prod. A	Prod. B	Prod. C	Sales	Evaluation
s_1	165	120	0	0	10	250	*
s_2	172.69	130	0.769	0	10	265.38	*
s_3	180.38	140	1.538	0	10	280.77	Good
s_4	141.13	140	3	3	4.92	272.92	Good
s_5	148.38	150	5	2	4.75	278.75	Good
s_6	139.13	150	5	3	3.58	279.58	*
s_7	188.08	150	2.308	0	10	296.15	*
s_8	159	150	6	0	6	270	*
s_9	140.5	150	6	2	3.67	271.67	Good
s_{10}	209.25	200	6	2	7.83	375.83	*
s_{11}	189.38	200	5	5	5.42	385.42	*
s_{12}	127.38	130	3	3	4.08	252.08	*
s_{13}	113.63	120	3	3	3.25	231.25	*

7. if Profit ≥ 127.38 and $x_A \geq 3$, then Time ≥ 130
($s_4, s_5, s_6, s_8, s_9, s_{10}, s_{11}, s_{12}$).
8. if Time ≤ 150 and $x_B \geq 2$, then Profit ≤ 148.38
($s_4, s_5, s_6, s_9, s_{12}, s_{13}$).
9. if $x_A \geq 3$ and $x_C \geq 4.08$, then Time ≥ 130
($s_4, s_5, s_8, s_{10}, s_{11}, s_{12}$).
10. if Sales ≥ 265.38 , then Time ≥ 130
($s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}$).

Then, the DM has been asked if he/she was satisfied with one of the proposed Pareto optimal solutions. Since his/her answer was negative, he/she was requested to indicate a subset of relatively “good” solutions which are indicated in the “Evaluation” column of Table 11.

Taking into account the ordinal classification of Pareto optimal solutions into “good” and “others,” made by the DM, 12 decision rules have been induced from the lower approximation of “good” solutions. The frequency of the presence of objectives in the premises of the rules gives a first idea of the importance of the considered objectives. These frequencies are the following:

- Profit: $\frac{4}{12}$
- Time: $\frac{12}{12}$
- Production of A : $\frac{7}{12}$

- Production of B : $\frac{4}{12}$
- Production of C : $\frac{5}{12}$
- Sales: $\frac{5}{12}$

The following potentially interesting decision rules were presented to the DM:

1. if Profit ≥ 140.5 and Time ≤ 150 and $x_B \geq 2$, then product mix is good (s_4, s_5, s_9).
2. if Time ≤ 140 and $x_A \geq 1.538$ and $x_C \geq 10$, then product mix is good (s_3).
3. if Time ≤ 150 and $x_B \geq 2$ and $x_C \geq 4.75$, then product mix is good (s_4, s_5).
4. if Time ≤ 140 and Sales ≥ 272.9167 , then product mix is good (s_3, s_4).
5. if Time ≤ 150 and $x_B \geq 2$ and $x_C \geq 3.67$ and Sales ≥ 271.67 , then product mix is good (s_4, s_5, s_9).

Among these decision rules, the DM has selected rule (1) as the most adequate to his/her preferences. This rule permits to define the following constraints reducing the feasible region of the production mix problem:

- $20x_A + 30x_B + 25x_C - (x_A + 2x_B + 0.75x_C)6 - (0.5x_A + x_B + 0.5x_C)8 \geq 140.5$
[Profit ≥ 140.5]

- $5x_A + 8x_B + 10x_C + 8x_A + 6x_B + 2x_C \leq 150$
[Time ≤ 150]
- $x_B \geq 2$ [Production of $B \geq 2$]

These constraints have been considered together with the original constraints for the production mix problem, and a new sample of representative Pareto optimal solutions shown in Table 12 have been calculated and presented to the DM, together with the following potentially interesting association rules:

- 1') if Time ≤ 140 , then Profit ≤ 174 and $x_C \leq 9.33$
and Sales ≤ 293.33
($s5', s6', s7', s8', s9', s10', s11', s12'$).
- 2') if $x_A \geq 2$, then $x_B \leq 3$ and Sales ≤ 300.83
($s2', s3', s4', s6', s7', s9'$).
- 3') if $x_A \geq 2$, then Profit ≤ 172 and $x_C \leq 8$
($s2', s3', s4', s6', s7', s9'$).
- 4') if Time ≤ 140 , then $x_A \leq 2$ and $x_B \leq 3$
($s5', s6', s7', s8', s9', s10', s11', s12'$).
- 5') if Profit ≥ 158.25 , then $x_A \leq 2$
($s1', s3', s4', s5', s6', s8'$).
- 6') if $x_A \geq 2$, then Time ≥ 130
($s2', s3', s4', s6', s7', s9'$).
- 7') if $x_C \geq 7.17$, then $x_A \leq 2$ and $x_B \leq 2$
($s1', s3', s5', s6', s8', s10'$).
- 8') if $x_C \geq 6$, then $x_A \leq 2$ and $x_B \leq 3$
($s1', s3', s4', s5', s6', s7', s8', s9', s10', s11', s12'$).

- 9') if $x_C \geq 7$, then Time ≥ 125 and $x_B \leq 2$
($s1', s3', s5', s6', s8', s10', s11'$).
- 10') if Sales ≥ 280 , then Time ≥ 140 and $x_B \leq 3$
($s1', s2', s3', s4', s5', s7'$).
- 11') if Sales ≥ 279.17 , then Time ≥ 140
($s1', s2', s3', s4', s5', s6', s7'$).
- 12') if Sales ≥ 272 , then Time ≥ 130
($s1', s2', s3', s4', s5', s6', s7', s8'$).

The DM has been asked again if he/she was satisfied with one of the proposed Pareto optimal solutions. Since his/her answer was negative, (s)he was requested again to indicate a subset of relatively “good” solutions, which are indicated in the “Evaluation” column of Table 12.

Taking into account the ordinal classification of Pareto optimal solutions into “good” and “others,” made by the DM, eight decision rules have been induced from the lower approximation of “good” solutions. The frequencies of the presence of objectives in the premises of the rules are the following:

- Profit: $\frac{2}{8}$
- Time: $\frac{1}{8}$
- Production of A : $\frac{5}{8}$
- Production of B : $\frac{3}{8}$
- Production of C : $\frac{3}{8}$
- Sales: $\frac{2}{8}$

Rough Sets in Decision-Making, Table 12 A sample of Pareto optimal solutions proposed in the second iteration

Solution	Profit	Time	Prod. A	Prod. B	Prod. C	Sales	Evaluation
$s1'$	186.53	150	0.154	2	10	313.08	*
$s2'$	154.88	150	3	3	5.75	293.75	*
$s3'$	172	150	2	2	8	300	Good
$s4'$	162.75	150	2	3	6.83	300.83	Good
$s5'$	174	140	0	2	9.33	293.33	*
$s6'$	158.25	140	2	2	7.17	279.17	Good
$s7'$	149	140	2	3	6	280	*
$s8'$	160.25	130	0	2	8.5	272	Good
$s9'$	144.5	130	2	2	6.33	258.33	*
$s10'$	153.38	125	0	2	8.08	262.08	*
$s11'$	145.5	125	1	2	7	255	Good
$s12'$	141.56	125	1.5	2	6.46	251.46	Good

The following potentially interesting decision rules were presented to the DM:

1. if $\text{Time} \leq 125$ and $x_A \geq 1$, then product mix is good
(s_{11}' , s_{12}').
2. if $x_A \geq 1$ and $x_C \geq 7$, then product mix is good
(s_3' , s_6' , s_{11}').
3. if $x_A \geq 1.5$ and $x_C \geq 6.46$, then product mix is good
(s_3' , s_4' , s_6' , s_{12}').
4. if $\text{Profit} \geq 158.25$ and $x_A \geq 2$, then product mix is good
(s_3' , s_4' , s_6').
5. if $x_A \geq 2$ and $\text{Sales} \geq 300$, then product mix is good
(s_3' , s_4').

Among these decision rules, the DM has selected rule (4) as the most adequate to his/her preferences. This rule permits to define the following constraints reducing the Pareto optimal set of the production mix problem:

- $20x_A + 30x_B + 25x_C - (x_A + 2x_B + 0.75x_C)6 - (0.5x_A + x_B + 0.5x_C)8 \geq 158.25$
[Profit ≥ 158.25]
- $x_A \geq 2$ [Production of $A \geq 2$]

Let us observe that the first constraint is just strengthening an analogous constraint introduced in the first iteration (Profit ≥ 140.5).

Considering the new set of constraints, a new sample of representative Pareto optimal solutions shown in Table 13 has been calculated and presented to the DM, together with the following potentially interesting association rules:

- 1'') if $\text{Time} \leq 145$, then $x_A \leq 2$ and $x_B \leq 2.74$ and $\text{Sales} \leq 290.2$
(s_2'' , s_3'' , s_4'').
- 2'') if $x_C \geq 6.92$, then $x_A \leq 3$ and $x_B \leq 2$ and $\text{Sales} \leq 292.92$
(s_3'' , s_4'' , s_5'').
- 3'') if $\text{Time} \leq 145$, then Profit ≤ 165.13 and $x_A \leq 2$ and $x_C \leq 7.58$
(s_2'' , s_3'' , s_4'').
- 4'') if $x_C \geq 6.72$, then $x_B \leq 2.74$
(s_2'' , s_3'' , s_4'' , s_5'').
- 5'') if $\text{Sales} \geq 289.58$, then Profit ≤ 165.13 and Time ≥ 145 and $x_C \leq 7.58$
(s_1'' , s_2'' , s_3'' , s_5'').

The DM has been asked again if he/she was satisfied with one of the presented Pareto optimal solutions shown in Table 13 and this time he/she declared that solution s_3'' is satisfactory for him/her. This ends the interactive procedure.

Characteristics of the IMO-DRSA

The interactive procedure presented in section "Interactive Multiobjective Optimization Using DRSA (IMO-DRSA)" can be analyzed from the point of view of input and output information. As to the input, the DM gives preference information by answering easy questions related to ordinal classification of some representative solutions into two classes ("good" and "others"). Very often, in multiple criteria decision analysis, in general, and in interactive multiobjective optimization, in particular, the preference information has to be given in terms of preference model parameters, such as importance weights, substitution rates, and various thresholds (see Fishburn (1967) for the multiple attribute utility theory and Brans et al. (2005),

Rough Sets in Decision-Making, Table 13 A sample of Pareto optimal solutions proposed in the third iteration

Solution	Profit	Time	Prod. A	Prod. B	Prod. C	Sales	Evaluation
s_1''	158.25	150	2	3.49	6.27	301.24	*
s_2''	158.25	145	2	2.74	6.72	290.20	*
s_3''	165.13	145	2	2	7.58	289.58	Selected
s_4''	158.25	140	2	2	7.17	279.17	*
s_5''	164.13	150	3	2	6.92	292.93	*
s_6''	158.25	145.3	3	2	6.56	284.02	*

Figueira et al. (2005), Martel and Matarazzo (2005), Roy and Bouyssou (1993) for outranking methods; for some well-known interactive multi-objective optimization methods requiring preference model parameters, see the Geoffrion-Dyer-Feinberg method (Geoffrion et al. 1972), the method of Zionts and Wallenius (1976, 1983), and the interactive surrogate worth tradeoff method (Chankong and Haimes 1978, 1983) requiring information in terms of marginal rates of substitution, the reference point method (Wierzbicki 1980) requiring a reference point and weights to formulate an achievement scalarizing function, the light beam search method (Jaszkiewicz and Słowiński 1999) requiring information in terms of weights and indifference, preference, and veto thresholds, being typical parameters of ELECTRE methods). Eliciting such information requires a significant cognitive effort on the part of the DM. It is generally acknowledged that people often prefer to make exemplary decisions and cannot always explain them in terms of specific parameters. For this reason, the idea of inferring preference models from exemplary decisions provided by the DM is very attractive. The output result of the analysis is the model of preferences in terms of “*if... then...*” decision rules which is used to reduce the Pareto optimal set iteratively, until the DM selects a satisfactory solution. The decision rule preference model is very convenient for decision support, because it gives argumentation for preferences in a logical form, which is intelligible for the DM, and identifies the Pareto optimal solutions supporting each particular decision rule. This is very useful for a critical revision of the original ordinal classification of representative solutions into the two classes of “good” and “others.” Indeed, decision rule preference model speaks the same language of the DM without any recourse to technical terms, like utility, tradeoffs, scalarizing functions, and so on.

All this implies that IMO-DRSA has a transparent feedback organized in a learning oriented perspective, which permits to consider this procedure as a “glass box,” contrary to the “black box” characteristic of many procedures giving final result without any clear explanation. Note that with the proposed procedure, the DM learns about the shape of the Pareto optimal set using the association rules.

They represent relationships between attainable values of objective functions on the Pareto optimal set in logical and very natural statements. The information given by association rules is as intelligible as the decision rule preference model, since they speak the language of the DM and permit him/her to identify the Pareto optimal solutions supporting each particular association rule.

Thus, decision rules and association rules give an explanation and a justification of the final decision, that does not result from a mechanical application of a certain technical method, but rather from a mature conclusion of a decision process based on active intervention of the DM.

Observe, finally, that the decision rules representing preferences and the association rules describing the Pareto optimal set are based on ordinal properties of objective functions only. Differently from methods involving some scalarization (almost all existing interactive methods), in any step the proposed procedure does not aggregate the objectives into a single value, avoiding operations (such as averaging, weighted sum, different types of distance, achievement scalarization) which are always arbitrary to some extent. Observe that one could use a method based on a scalarization to generate the representative set of Pareto optimal solutions, nevertheless, the decision rule approach would continue to be based on ordinal properties of objective functions only, because the dialogue stage of the method operates on ordinal comparisons only. In the proposed method, the DM gets clear arguments for his/her decision in terms of “*if... then...*” decision rules and the verification if a proposed solution satisfies these decision rules is particularly easy. This is not the case of interactive multiobjective optimization methods based on scalarization. For example, in the methods using an achievement scalarization function, it is not evident what does it mean for a solution to be “close” to the reference point. How to choose a suitable distance? How to justify the choice of the weights used in the achievement function? What is their interpretation? Observe, instead, that the method proposed in this entry operates on data using ordinal comparisons which would not be affected by any increasing monotonic transformation of scales, and this

ensures the meaningfulness of results from the point of view of measurement theory (see, e.g., Roberts 1979).

With respect to computational aspects of the method, notice that the decision rules can be calculated efficiently in few seconds only using the algorithms presented in Greco et al. (2001c, (2002c). When the number of objective functions is not too large to be effectively controlled by the DM (let us say seven plus or minus two, as suggested by Miller (1956)), then the decision rules can be calculated in a fraction of one second. In any case, the computational effort grows exponentially with the number of objective functions, but not with respect to the number of considered Pareto optimal solutions, which can increase with no particularly negative consequence on calculation time.

Some Dedicated Applications of the IMO-DRSA: Financial Portfolio and Project Portfolio

Two noteworthy applications of IMO-DRSA concern financial portfolio selection (Greco et al. 2013) and project portfolio decision problem (Barbati et al. 2018). We shall present both of them in the two following subsections.

IMO-DRSA for financial portfolio

A financial portfolio is a collection of securities that are tradable financial assets. A classical problem in finance is how to compose a portfolio in order to maximize the expected return and to minimize the risk. This is in fact the Markowitz's celebrated mean-variance portfolio optimization problem (Markowitz 1952) which can be formalized as follows. Assuming set $J = \{1, \dots, j, \dots, z\}$ of risk securities (assets) available on the capital market, a portfolio $\mathbf{x} \in \mathcal{R}^z$ is described by an allocation vector $\mathbf{x} = [x_1, \dots, x_z]$, with $\sum_{j=1}^z x_j = 1$. If one knew all possible outcomes and corresponding probabilities of each security $j \in J$, then one could compute easily the expected value $E(R_j)$ of return R_j of security j . Let us denote by $R(\mathbf{x})$ the return of portfolio $\mathbf{x}$, i.e.,

$$R(\mathbf{x}) = \sum_{j=1}^z x_j R_j.$$

The expected value of portfolio return $R(\mathbf{x})$ is therefore $E(R(\mathbf{x})) = \sum_{j=1}^z x_j E(R_j)$. The risk can

be expressed in terms of variance $\sigma^2(R(\mathbf{x}))$ defined as expected squared deviation of portfolio return $R(\mathbf{x})$ from its expected return $E(R(\mathbf{x}))$, i.e., $\sigma^2(R(\mathbf{x})) = E[(R(\mathbf{x}) - E(R(\mathbf{x})))^2]$. Under the hypothesis of normally distributed returns of risk securities, the variance $\sigma^2(R(\mathbf{x}))$ becomes:

$$\sigma^2(R(\mathbf{x})) = \sum_{j=1}^z x_j^2 \sigma_j^2 + 2x_j x_h \sum_{j=1}^z \sum_{h>j}^z \sigma_{jh},$$

where σ_j^2 is the variance of return of risk security $j \in J$, and σ_{jh} is the covariance of returns of risk securities $j, h \in J$. The choice of variance as a risk measure is due to the easy computation of variance, and because the mean and the variance of return contain all relevant information about the distribution if returns are normally distributed. This is why this hypothesis is usually assumed in portfolio optimization. As a substitute of variance, one is often using its square root $\sigma(R(\mathbf{x}))$, called standard deviation.

Thus, according to the Markowitz model, the portfolio selection problem is a bi-objective optimization problem, where the two objectives are: maximization of the expected return $E(R(\mathbf{x}))$ and minimization of the variance $\sigma^2(R(\mathbf{x}))$. In such a context, the efficient frontier is composed of portfolios $\mathbf{x}$ for which it is not possible to increase the expected return $E(R(\mathbf{x}))$ without increasing the variance $\sigma^2(R(\mathbf{x}))$, or, equivalently, it is not possible to decrease the variance $\sigma^2(R(\mathbf{x}))$ without decreasing the expected return $E(R(\mathbf{x}))$. For a given value of variance α , a portfolio $\mathbf{x}$ is in the efficient frontier if it maximizes the expected return $E(R(\mathbf{x}))$, i.e., if $\mathbf{x}$ is a solution of the following optimization problem (P1):

$$\text{maximize } E(R(\mathbf{x}))$$

under the constraints

$$\begin{aligned} \sigma^2(R(\mathbf{x})) &= \alpha, \\ \sum_{j=1}^z x_j &= 1. \end{aligned}$$

Even if the Markowitz model is of fundamental importance for finance theory, it presents several

weak points, the most important of them are the following:

- The mean-variance model of Markowitz fails to satisfy monotonicity property, such that there can be a situation where the DM adopting the mean-variance approach can prefer investment a having a prospect dominated by a prospect of another investment b , i.e., the DM selects a even if it gives a not better return than b in each state of the world (see, e.g., Bigelow 1993).
- The hypothesis of normal distributions of returns, on which the Markowitz approach is largely based, has been put in doubt by the work of Mandelbrot (1963) and Fama (1965), who proposed the stable Pareto distribution as a statistical model for an asset return.
- The variance presents several problems as a risk measure, the first of which is that it aggregates negative deviations from the average that are not desirable, with positive deviations from the average which, instead, are desirable; for this reason, Markowitz (1959) proposed to consider the semi-variance based on negative deviations only. In fact, recently, many alternative risk measures have been considered. The most well known is the so-called value at risk (VaR) (Jorion 2006), that is the worst $\alpha\%$ quantile of return, and the conditional value at risk (CVaR) (also known as average value at risk (AVaR) and expected tail loss (ETL)) (Rockafellar and Uryasev 2000) being the expected loss in the worst $\alpha\%$ quantile of return.

The IMO-DRSA approach to the Markowitz portfolio problem, formulated as a multiobjective optimization problem where a limited number of quantiles on the distribution of return is maximized, overcomes all three weak points of the Markowitz model discussed above. Precisely, let us observe that:

- With respect to the monotonicity, IMO-DRSA is safe because if an investment a is dominating another investment b , in the sense that it gives a not smaller return in all states of the world, then

a dominates b also with respect to any subset of considered quantiles.

- With respect to non-normal distribution of returns, IMO-DRSA can work with all kinds of distributions, because it only requires the estimation of a small number of meaningful quantiles of the portfolio return distribution.
- With respect to risk measures based on quantiles, so much appreciated in the current practice of finance, IMO-DRSA is perfectly in line because it takes into account only a limited number of quantiles (7 ± 2) (Miller 1956).

The IMO-DRSA approach to the selection of financial portfolios is based on the following alternance of computation and dialog with the DM:

- In a computation phase, a small sample of financial portfolios non-dominated with respect to the selected quantiles is computed.
- In a dialog phase, the DM classifies some financial portfolios from the current sample as (relatively) “good.”
- In the next computation phase, decision rules characterizing the “good” portfolios in terms of considered quantiles are induced by DRSA.
- In the next dialog phase, the DM selects some decision rules that in her/his opinion represent the best her/his preferences.
- In the following computation phase, the selected rules are transformed into constraints added to the multiobjective optimization problem, so that the values of quantiles specified in these rules become thresholds on quantiles of return to be attained by calculated portfolios; then a new sample of non-dominated portfolios is calculated and presented to the DM, and the procedure iterates.

The rules presented to the DM have a well understandable syntax, such as: if $\alpha_1\%$ quantile of return is not smaller than $r_1\%$, and $\alpha_2\%$ quantile of return is not smaller than $r_2\%$, then the financial portfolio $\mathbf{x}$ is “good.”

IMO-DRSA for project portfolio In the project portfolio decision problem a set of projects

constituting the portfolio has to be selected so as to optimize some objectives (criteria) subject to constraints related to limited resources (e.g., available budget). The usual approach to such a problem is the maximization of multiple attribute values corresponding to the considered objectives, subject to some resource constraints (Salo et al. 2011). The main drawback of this approach is that the selected portfolio of projects can have very good values on a small set of criteria and rather bad values on the remaining criteria. To explain this situation Barbati et al. (2018) gives the following example. Consider the manager of a football club that have to select 16 players, 11 regular ones and 5 substitutes. Let us suppose that the best 16 football players are all goalkeepers and the budget permits to recruit all of them. In this case, maximization of the multiple attribute values of the 16 players would lead to a team with 16 goalkeepers, which is clearly paradoxical since, clearly, a football team needs to involve defenders, midfielders, and forwards. In the IMO-DRSA approach proposed by Barbati et al. (2018), the portfolio decision problem is handled by maximizing the number of projects attaining some meaningful levels of quality on criteria at hand. This permits to focus the search on portfolios with projects having good evaluations well distributed over the considered criteria, taking also into account preferences of the DM. Coming back to the above example, the manager would take care that a certain number of players would have good performances on ball handling, positioning, and quick reflexes, that is, the typical criteria for goalkeepers, but without forgetting to recruit a sensible number of players being good with respect to passing/receiving, dribbling, shooting, tackling, trapping, sprinting, endurance, heading, and juggling. The final result would then be a team composed of a reasonable number of defenders, midfielders, and forwards, apart from the two or three goalkeepers.

Formally, the IMO-DRSA approach to the project portfolio decision problem is formulated as follows. Let us consider a portfolio decision problem with a set of projects $A = \{a_1, \dots, a_j, \dots, a_n\}$, evaluated on a set of criteria $G = \{g_1, \dots, g_i, \dots, g_h\}$, supposed (without the loss of generality) to be of a gain type, that is the greater $g_i(a_j)$, the

better value of project $a_j \in A$ on criterion g_i . The cost of each project $a_j \in A$ is $c_j \in \mathfrak{R}^+$, and the available budget is C .

For each $g_i \in G$, the DM fixes a set L_i consisting of $G(i)$ quality thresholds

$$L_i = \{l_{1,i}, \dots, l_{G(i),i} : l_{1,i} < l_{2,i} < \dots < l_{G(i),i}\}, \quad (1)$$

permitting to define a set C_i composed of $G(i) + 1$ qualitative satisfaction classes $C_{t,i}$,

$$C_i = \{C_{1,i}, \dots, C_{G(i)+1,i}\} \quad (2)$$

such that the greater $t = 1, \dots, G(i) + 1$, the more preferred is the project from class $C_{t,i}$. Projects $a_j \in A$ are assigned to satisfaction classes $C_{t,i} \in C_i$ according to the following rule: for all $a_j \in A$

- a_j is assigned to class $C_{1,i}$ if $g_i(a_j) < l_{1,i}$.
- a_j is assigned to class $C_{t,i}$ if $l_{t-1,i} \leq g_i(a_j) < l_{t,i}$.
- a_j is assigned to class $C_{G(i)+1,i}$ if $l_{G(i),i} \leq g_i(a_j)$.

For example, when considering criterion g_i with a numerical evaluation scale $[0, 100]$, one can fix a set $L_i = \{15, 45, 80\}$ of $G(i) = 3$ quality thresholds to define four satisfaction classes: “weakly satisfactory,” “satisfactory,” “very satisfactory,” and “extremely satisfactory.” Then, for any project $a_j \in A$, we have that with respect to criterion g_i :

- a_j is “weakly satisfactory” if $g_i(a_j) < 15$.
- a_j is “satisfactory” if $15 \leq g_i(a_j) < 45$.
- a_j is “very satisfactory” if $45 \leq g_i(a_j) < 80$.
- a_j is “extremely satisfactory” if $80 \leq g_i(a_j)$.

Each subset of projects $P \subseteq A$ is a potential portfolio that can be identified with a vector $\mathbf{x} = [x_1, \dots, x_j, \dots, x_n]$, such that for all $a_j \in A$:

$$x_j = \begin{cases} 1, & \text{if } a_j \in A \text{ is contained in } P, \\ 0, & \text{otherwise.} \end{cases}$$

For each portfolio $P \subseteq A$, each criterion $g_i \in G$ and each satisfaction level $l_{t,i} \in L_i$, we consider the set of projects attaining threshold $l_{t,i}$:

$$P_{t,i} = \{a_j \in P : g_i(a_j) \geq l_{t,i}\}. \quad (3)$$

For the sake of simplicity, in what follows we shall write $|P_{t,i}|$, as $F_{t,i}(\mathbf{x})$ to refer to portfolio P identified by vector $\mathbf{x}$, $g_i \in G$ and $l_{t,i} \in L_i$.

Taking into account all above elements, we formulate the portfolio decision problem as the following multiobjective optimization problem:

$$\max F_{t,i}(\mathbf{x}), \quad \forall g_i \in G, \quad \forall l_{t,i} \in L_i \quad (4)$$

subject to:

$$\sum_{a_j \in A} c_j x_j \leq C \quad (5)$$

$$x_j = \begin{cases} 1, & \text{if } a_j \in A \text{ is contained in } P \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

$$j = 1, \dots, n.$$

Clearly, the above problem is a multiobjective 0–1 linear programming problem with $\sum_{i=1}^h G(i)$ objectives, that can be handled with some of the many algorithms, mainly exact, provided in the literature (for a review, see Ehrgott et al. 2016).

Using the IMO-DRSA approach to the project portfolio selection, phases of computation and dialog with the DM alternate in the same way as in the case of financial portfolio described in the previous subsection.

Also in this case, the rules presented to the DM are well understandable for the DM. For example: if in a portfolio there are at least 5 projects being at least satisfactory on criterion g_i , and at least 7 projects being at least very satisfactory on criterion g_i' , then the portfolio is “good.”

An Abstract Algebra for Reasoning About Multiple Criteria Decisions: The Bipolar Pawlak-Brouwer-Zadeh Lattice

Several algebraic structures have been proposed to give a formal representation of classical rough set theory (see, e.g., Chap. 12 in Polkowski 2002). In particular, among the rough set algebraic structures, there is the Brouwer-Zadeh lattice (Cattaneo

1997). The Brouwer-Zadeh lattice has been extended to a bipolar Brouwer-Zadeh lattice (Greco et al. 2012b), giving an algebraic model to dominance-based rough set. Moreover, in Greco et al. (2012a), another extension of the Brouwer-Zadeh lattice has been proposed, called Pawlak-Brouwer-Zadeh lattice. In this extension, the basic vagueness is represented by a pair (A, B) , with $A \cap B = \emptyset$, A being the necessity kernel and B being the non-possibility kernel, and ambiguity due to the coarseness of information is introduced through a new operator, called Pawlak operator. Pawlak operator assigns a pair (C, D) to the pair (A, B) , such that C and D represent the lower approximations of A and B , respectively. Finally, very recently, the bipolar Pawlak-Brouwer-Zadeh lattice has been proposed in Greco et al. (2017) to get an algebraic model permitting to distinguish vagueness from ambiguity in DRSA. Since the operators of the bipolar Pawlak-Brouwer-Zadeh lattice constitute a powerful tool for reasoning about multiple criteria decisions, we introduce its core idea and show a simple didactic example illustrating its utility.

In a given universe of objects of interest U , for example, a set of companies evaluated on a set P of criteria, for which a risk of bankruptcy have to be assessed, let us consider a family 3^U of all pairs of disjoint subsets of objects (A, B) , i.e., $A, B \subseteq U$ and $A \cap B = \emptyset$. (A, B) can be seen as a result of an expert's decision who assigned to set A those companies from U which are safe, and to set B those which are risky, so that $U - A - B$ is the set of doubtful companies for which the expert is not able to give a univocal assessment as safe or risky. In this framework, some meaningful operations can be defined on 3^U . To give an idea, let us consider the assessments of two experts represented by the pairs $(A^1, B^1), (A^2, B^2) \in 3^U$. A first operation on 3^U is the join $\sqcup$ for which $(A^1, B^1) \sqcup (A^2, B^2) = (A^1 \cup A^2, B^1 \cap B^2)$, that is the join $\sqcup$ considers companies as safe when they are safe for at least one expert, and risky when they are risky for both experts. A second operation on 3^U is the meet $\sqcap$ for which $(A^1, B^1) \sqcap (A^2, B^2) = (A^1 \cap A^2, B^1 \cup B^2)$, that is the meet $\sqcap$ considers companies as safe when they are safe for both experts, and risky when they are risky for at least

one expert. Another operation on 3^U is the Kleene negation – for which, for all $(A, B) \in 3^U$, $(A, B)^- = (B, A)$, that is, the Kleene negation – makes of safe companies risky ones and vice versa. A further operation is the Brouwer negation $\approx$, for which, for all $(A, B) \in 3^U$, $(A, B)^\approx = (B, U - B)$, that is, the Brouwer negation $\approx$ makes of risky companies safe ones, and of all the others risky ones. The last operation on 3^U is N being the Pawlak operator exploiting knowledge about dominance relation in U , for which, for all $(A, B) \in 3^U$, $(A, B)^N = (\underline{P}(A^\geq), \underline{P}(B^\leq))$, that is the Pawlak operator N transforms the set of safe companies A in its upward rough lower approximation $\underline{P}(A^\geq)$, and the set of risky companies B in its downward rough lower approximation $\underline{P}(B^\leq)$. Using the above five elementary operations $\sqcup, \sqcap, ^-, \approx, ^N$, other interesting composite operations can be defined, so that bipolar Pawlak-Brouwer-Zadeh lattice becomes a grammar for reasoning about multiple criteria decision-making.

Conclusions

Rough set theory is a mathematical tool for dealing with granularity of information and possible inconsistencies in the description of objects. Considering this description as an input data about a decision problem, the rough set approach permits to structure this description into lower and upper approximations, corresponding to certain and possible knowledge about the problem. Induction algorithms run on these approximations discover, in turn, certain and possible decision rules that facilitate an understanding of the DM's preferences, and that enable a recommendation concordant with these preferences.

The original version of the rough set approach, based on indiscernibility or similarity relation, and typical induction algorithms considered within machine learning, data mining and knowledge discovery, deal with data describing problems of taxonomy-type classification, i.e., problems where neither the attributes describing the objects, nor the classes to which the objects are

assigned, are ordered. On the other hand, multiple criteria decision-making deals with problems where descriptions (evaluations) of objects by means of attributes (criteria), as well as decisions in classification, choice, and ranking problems, are ordered. Moreover, in data describing multiple criteria decision-making, there exists a monotonic relationship between conditions and decisions, like “the bigger the house, the more expensive it is.” The generalization of the rough set approach and of the induction algorithms about problems in which order properties and monotonic relationships are important gave birth to the dominance-based rough set approach (DRSA) which made a breakthrough in scientific decision support.

The main features of DRSA are the following:

- Preference information necessary to deal with any multiple criteria decision problem, or with decision under uncertainty, is asked to the DM just in terms of exemplary decisions.
- The rough set analysis of preference information supplies some useful elements of knowledge about the decision situation; these are: the relevance of attributes or criteria, the minimal subsets of attributes or criteria (reducts) conveying the relevant knowledge contained in the exemplary decisions, and the set of indispensable attributes or criteria (core).
- DRSA can deal with preference information concerning taxonomy-type classification, ordinal classification, choice, ranking, multi-objective optimization, and decision under uncertainty.
- The preference model induced from preference information structured by DRSA is expressed in a natural and comprehensible language of “if... then...” decision rules.
- Suitable procedures have been proposed to exploit the results of application of the decision rule preference model on a set of objects or pairs of objects in order to workout a recommendation.
- No prior discretization of quantitative condition attributes or criteria is necessary.
- Heterogeneous information (qualitative and quantitative, ordered and non-ordered, nominal and ordinal, quantitative and numerical

nonquantitative scales of preferences) can be processed within DRSA.

- The proposed methodology fulfils some desirable properties for both rough set approach (the approximated sets include lower approximation and are included in upper approximation, and the complementarity property is satisfied) and for multiple criteria decision analysis (the decision rule preference model is formally equivalent to the nonadditive, non-transitive, and non-complete conjoint measurement model, and to a more general model for preferences defined on all kinds of scales).
- The decision rule preference model resulting from DRSA is more general than all existing models of conjoint measurement, due to its capacity of handling inconsistent preferences (a new model of conjoint measurement is formally equivalent to the decision rule preference model handling inconsistencies).
- The decision rule preference model fulfils the postulate of transparency and interpretability of preference models in decision support; each decision rule can be clearly identified with those parts of the preference information (decision examples) which support the rule; the rules inform the DM in a quasi-natural language about the relationships between conditions and decisions; in this way, the rules permit traceability of the decision support process and give understandable justifications for the decision to be made.
- The proposed methodology is based on elementary concepts and mathematical tools (sets and set operations, binary relations), without recourse to any algebraic or analytical structures.

The decision rules entering the preference model have a special syntax which involves partial evaluation profiles and dominance relations on these profiles. The traditional preference models, which are the utility function and the outranking relation, can be represented in terms of equivalent decision rules. The clarity of the rule representation of preferences enables one to see the limits of these aggregation functions. In several studies (Greco

et al. 2002b, 2004a, 2016b; Słowiński et al. 2002b) there was proposed an axiomatic characterization of all three kinds of preference models in terms of conjoint measurement theory and in terms of a set of decision rules. The given axioms of “cancellation property” type are the weakest possible. In comparison to other studies on the characterization of aggregation functions, these axioms do not require any preliminary assumptions about the scales of criteria. A side result of these investigations is that the decision rule preference model is the most general among the known aggregation functions.

Prospective Directions of Applications

The entry shows that the dominance-based rough set approach is a very powerful tool for decision analysis and support. Its potential goes, however, beyond the theoretical frame considered in this entry. There are many possibilities of applying DRSA to real life problems. The non-exhaustive list of potential applications includes:

- Decision support in medicine: In this area, there are already many interesting applications based on the classical rough set approach (see, e.g., Michalowski et al. 2003, 2005; Pawlak et al. 1986), as well as on DRSA, where ordered value sets of medical attributes, and monotonic relationships between values of these attributes and the degree of gravity of a disease are taken into account (see, e.g., Blasco et al. 2015; Gowin et al. 2017; Wilk et al. 2005).
- Customer satisfaction analysis: Theoretical foundations for application of DRSA in this field are available in Greco et al. (2007a); customer satisfaction can be improved by interventions recommended by dominance-based rules; Greco et al. (2005) reports a study of measurable effects of such interventions.
- Bankruptcy risk evaluation: This is a field of many potential applications, as can be seen from promising results reported, e.g., in Greco et al. (1998a), Słowiński and Zopounidis (1995), Słowiński et al. (1997).

- Operational research problems, such as location, routing, scheduling, or inventory management: these are problems formulated either in terms of classification of feasible solutions (see, e.g., Gorsevski and Jankowski 2008), or in terms of interactive multiobjective optimization, for which the IMO-DRSA (Greco et al. 2008a) procedure is suitable.
- Finance: This is a domain where DRSA for decision under uncertainty has to be combined with interactive multiobjective optimization using IMO-DRSA; successful applications of IMO-DRSA to portfolio selection problems have been presented in Barbati et al. (2018), Greco et al. (2013).
- Chemoinformatics: Decision rules induced by DRSA from data describing activity of chemical compounds provide indications for synthesis of new compounds with better properties (Gardiner and Gillet 2015; Pałkowski et al. 2014a, b, 2015).
- Sustainable development: DRSA has shown its utility in discovering rules of cleaner production from data describing technological processes (Cinelli et al. 2015, 2017).
- Ecology: Assessment of the impact of human activity on the ecosystem is a challenging problem for which the presented methodology has appeared to be suitable (Flinkman et al. 2000; Rossi et al. 1999).

Bibliography

- Barbati M, Greco S, Słowiński R (2018) Optimization of multiple satisfaction levels in portfolio decision analysis. *Omega* 78:192–204
- Bigelow JP (1993) Consistency of mean-variance analysis and expected utility analysis: a complete characterization. *Econ Lett* 43:187–192
- Blasco H, Błaszczyński J, Billaut J-C, Nadal L, Pradat P-F, Devos D, Moreau C, Andres CR, Emond P, Corcia P, Słowiński R (2015) Comparative analysis of targeted metabolomics: dominance-based rough set approach versus orthogonal partial least square-discriminant analysis. *J Biomed Inform* 53:291–299
- Błaszczyński J, Słowiński R, Szeląg M (2011) Sequential covering rule induction algorithm for variable consistency rough set approaches. *Inf Sci* 181:987–1000
- Brans JP, Mareschal B (2005) Promethee methods. Chapter 5. In: Figueira J, Greco S, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, Berlin, pp 163–195
- Cattaneo G (1997) Generalized rough sets (Preclusivity fuzzy-intuitionistic (BZ) lattices). *Stud Logica* 58: 47–77
- Chankong V, Haimes YY (1978) The interactive surrogate worth trade-off (ISWT) method for multiobjective decision-making. In: Zionts S (ed) Multiple criteria problem solving. Springer, Berlin/New York, pp 42–67
- Chankong V, Haimes YY (1983) Multiobjective decision making theory and methodology. Elsevier Science Publishing Co., New York
- Cinelli M, Coles SR, Nadagouda MN, Błaszczyński J, Słowiński R, Varma RS, Kirwan K (2015) A green chemistry-based classification model for the synthesis of silver nanoparticles. *Green Chem* 17:2825–2839
- Cinelli M, Coles SR, Nadagouda MN, Błaszczyński J, Słowiński R, Varma RS, Kirwan K (2017) Robustness analysis of a green chemistry-based model for the classification of silver nanoparticles synthesis processes. *J Clean Prod* 162:938–948
- Corrente S, Greco S, Kadziński M, Słowiński R (2013) Robust ordinal regression in preference learning and ranking. *Mach Learn* 93:381–422
- Ehrgott M, Gandibleux X, Przybylski A (2016) Exact methods for multi-objective combinatorial optimisation. Chapter 19. In: Greco S, Figueira J, Ehrgott M (eds) Multiple criteria decision analysis: state-of-the-art surveys, 2nd edn. Springer, New York, pp 817–850
- Fama E (1965) Portfolio analysis in a stable Paretian market. *Manag Sci* 11:404–419
- Figueira J, Mousseau V, Roy B (2005) ELECTRE methods. Chapter 4. In: Figueira J, Greco S, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, Berlin, pp 133–162
- Fishburn PC (1967) Methods of estimating additive utilities. *Manag Sci* 13:435–453
- Flinkman M, Michalowski W, Nilsson S, Słowiński R, Susmaga R, Wilk S (2000) Use of rough sets analysis to classify Siberian forest ecosystem according to net primary production of phytomass. *Infor* 38:145–161
- Fortemps P, Greco S, Słowiński R (2008) Multicriteria decision support using rules that represent rough-graded preference relations. *Eur J Oper Res* 188: 206–223
- Gardiner EJ, Gillet VJ (2015) Perspectives on knowledge discovery algorithms recently introduced in Chemoinformatics: rough set theory, association rule mining, emerging patterns, and formal concept analysis. *J Chem Inf Model* 55:1781–1803
- Geoffrion A, Dyer J, Feinberg A (1972) An interactive approach for multi-criterion optimization, with an application to the operation of an academic department. *Manag Sci* 19:357–368
- Giove S, Greco S, Matarazzo B, Słowiński R (2002) Variable consistency monotonic decision trees. In: Alpigini JJ, Peters JF, Skowron A, Zhong N (eds) Rough sets

- and current trends in computing, LNAI 2475. Springer, Berlin, pp 247–254
- Gorsevski PV, Jankowski P (2008) Discerning landslide susceptibility using rough sets. *Comput Environ Urban Syst* 32:53–65
- Gowin E, Januszkievicz-Lewandowska D, Słowiński R, Błaszczyński J, Michalak M, Wysocki J (2017) With a little help from a computer: discriminating between bacterial and viral meningitis based on dominance-based rough set approach analysis. *Medicine* 96(32): e7635
- Greco S, Matarazzo B, Słowiński R (1998a) A new rough set approach to evaluation of bankruptcy risk. In: Zopounidis C (ed) *Operational tools in the Management of Financial Risks*. Kluwer, Dordrecht, pp 121–136
- Greco S, Matarazzo B, Słowiński R, Tsoukias A (1998b) Exploitation of a rough approximation of the outranking relation in multicriteria choice and ranking. In: Stewart TJ, van den Honert RC (eds) *Trends in multicriteria decision making*. LNEMS 465. Springer, Berlin, pp 45–60
- Greco S, Matarazzo B, Słowiński R (1999a) The use of rough sets and fuzzy sets in MCDM. Chapter 14. In: Gal T, Stewart T, Hanne T (eds) *Advances in multiple criteria decision making*. Kluwer, Boston, pp 14.1–14.59
- Greco S, Matarazzo B, Słowiński R (1999b) Rough approximation of a preference relation by dominance relations. *Eur J Oper Res* 117:63–83
- Greco S, Matarazzo B, Słowiński R (2000) Extension of the rough set approach to multicriteria decision support. *Infor* 38:161–196
- Greco S, Matarazzo B, Słowiński R (2001a) Rough sets theory for multicriteria decision analysis. *Eur J Oper Res* 129:1–47
- Greco S, Matarazzo B, Słowiński R (2001b) Rough set approach to decisions under risk. In: Ziarko W, Yao Y (eds) *Rough sets and current trends in computing*, LNAI 2005. Springer, Berlin, pp 160–169
- Greco S, Matarazzo B, Słowiński R, Stefanowski J (2001c) An algorithm for induction of decision rules consistent with dominance principle. In: Ziarko W, Yao Y (eds) *Rough sets and current trends in computing*, LNAI 2005. Springer, Berlin, pp 304–313
- Greco S, Matarazzo B, Słowiński R (2002a) Multicriteria classification, chapter 16.1.9. In: Kloesgen W, Zytkow J (eds) *Handbook of data mining and knowledge discovery*. Oxford University Press, pp 318–328
- Greco S, Matarazzo B, Słowiński R (2002b) Preference representation by means of conjoint measurement & decision rule model. In: Bouyssou D, Jacquet-Lagrèze E, Perny P, Słowiński R, Vanderpooten D, Vincke P (eds) *Aiding decisions with multiple criteria—essays in honor of Bernard Roy*. Kluwer, Dordrecht, pp 263–313
- Greco S, Matarazzo B, Słowiński R, Stefanowski J (2002c) Mining association rules in preference-ordered data. In: Hacid M-S, Ras ZW, Zighed DA, Kodratoff Y (eds) *Foundations of intelligent systems*, LNAI 2366. Springer, Berlin, pp 442–450
- Greco S, Matarazzo B, Słowiński R (2004a) Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *Eur J Oper Res* 158:271–292
- Greco S, Matarazzo B, Słowiński R (2004b) Dominance-based rough set approach to knowledge discovery, (I) – general perspective, (II) – extensions and applications. Chapters 20 and 21. In: Zhong N, Liu J (eds) *Intelligent technologies for information analysis*. Springer, Berlin, pp 513–612
- Greco S, Pawlak Z, Słowiński R (2004c) Can Bayesian confirmation measures be useful for rough set decision rules? *Eng Appl Artif Intell* 17:345–361
- Greco S, Matarazzo B, Pappalardo N, Słowiński R (2005) Measuring expected effects of interventions based on decision rules. *J Exp Theor Artif Intell* 17:103–118
- Greco S, Matarazzo B, Słowiński R (2006a) Dominance-based rough set approach to decision involving multiple decision makers. In: Greco S, Hata Y, Hirano S, Inuiguchi M, Miyamoto S, Nguyen HS, Słowiński R (eds) *Rough sets and current trends in computing (RSCTC 2006)*. LNAI 4259. Springer, Berlin, pp 306–317
- Greco S, Matarazzo B, Słowiński R (2006b) Dominance-based rough set approach to case-based reasoning. In: Torra V, Narukawa Y, Valls A, Domingo-Ferrer J (eds) *Modelling decisions for artificial intelligence*. LNAI 3885. Springer, Berlin, pp 7–18
- Greco S, Matarazzo B, Słowiński R (2007a) Customer satisfaction analysis based on rough set approach. *Z Betriebswirt* 16:325–339
- Greco S, Matarazzo B, Słowiński R (2007b) Dominance-based rough set approach as a proper way of handling graduality in rough set theory. In: *Transactions on rough sets VII*, LNCS 4400. Springer, Berlin, pp 36–52
- Greco S, Matarazzo B, Słowiński R (2008a) Dominance-based rough set approach to interactive multiobjective optimization. Chapter 5. In: Branke J, Deb K, Miettinen K, Słowiński (eds) *Multiobjective optimization: interactive and evolutionary approaches*. Springer, Berlin, pp 121–156
- Greco S, Mousseau V, Słowiński R (2008b) Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *Eur J Oper Res* 191:416–436
- Greco S, Matarazzo B, Słowiński R (2010a) Dominance-based rough set approach to decision under uncertainty and time preference. *Ann Oper Res* 176:41–75
- Greco S, Mousseau V, Słowiński R (2010b) Multiple criteria sorting with a set of additive value functions. *Eur J Oper Res* 207:1455–1470
- Greco S, Matarazzo B, Słowiński R (2012a) Distinguishing vagueness from ambiguity by means of Pawlak-Brouwer-Zadeh lattices. In: Greco S et al (eds) *International conference on information processing and Management of Uncertainty in knowledge-based systems*. Springer, Berlin, pp 624–632

- Greco S, Matarazzo B, Słowiński R (2012b) The bipolar complemented de Morgan Brouwer-Zadeh distributive lattice as an algebraic structure for the dominance-based rough set approach. *Fund Inform* 115:25–56
- Greco S, Matarazzo B, Słowiński R (2013) Beyond Markowitz with multiple criteria decision aiding. *J Bus Econ* 83:29–60
- Greco S, Figueira J, Ehrgott M (eds) (2016a) Multiple criteria decision analysis: state-of-the-art surveys, 2nd edn. Springer, New York
- Greco S, Matarazzo B, Słowiński R (2016b) Decision rule approach. Chapter 13. In: Greco S, Ehrgott M, Figueira J (eds) Multiple criteria decision analysis: state of the art surveys, 2nd edn, OR & MS 233. Springer, New York, pp 497–552
- Greco S, Słowiński R, Szczęch I (2016c) Measures of rule interestingness in four perspectives of confirmation. *Inf Sci* 346–347:216–235
- Greco S, Matarazzo B, Słowiński R (2017) Distinguishing vagueness from ambiguity in dominance-based rough set approach by means of a bipolar Pawlak-Brouwer-Zadeh lattice. In: Polkowski L et al (eds) International joint conference on rough sets, IJCRS 2017, Part II, LNAI 10314. Springer, Berlin, pp 81–93
- Jaskiewicz A, Słowiński R (1999) The “light beam search” approach – an overview of methodology and applications. *Eur J Oper Res* 113:300–314
- Jorion P (2006) Value at risk: the new benchmark for managing financial risk, 3rd edn. McGraw Hill, New York
- Kadziński M, Greco S, Słowiński R (2014) Robust ordinal regression for dominance-based rough set approach to multiple criteria sorting. *Inf Sci* 283:211–228
- Kadziński M, Słowiński R, Greco S (2015) Multiple criteria ranking and choice with all compatible minimal cover sets of decision rules. *Knowl-Based Syst* 89: 569–583
- Kadziński M, Słowiński R, Greco S (2016) Robustness analysis for decision under uncertainty with rule-based preference model. *Inf Sci* 328:321–339
- Mandelbrot B (1963) The variation of certain speculative prices. *J Bus* 36:394–419
- Markowitz HM (1952) Portfolio selection. *J Financ* 7: 77–91
- Markowitz HM (1959) Portfolio selection: efficient diversification of investments. Wiley, New York
- Martel JM, Matarazzo B (2005) Other outranking approaches. Chapter 6. In: Figueira J, Greco S, Ehrgott M (eds) Multiple criteria decision analysis: state of the art surveys. Springer, Berlin, pp 197–262
- Michalowski W, Rubin S, Słowiński R, Wilk S (2003) Mobile clinical support system for pediatric emergencies. *J Decis Support Syst* 36:161–176
- Michalowski W, Wilk S, Farion K, Pike J, Rubin S, Słowiński R (2005) Development of a decision algorithm to support emergency triage of scrotal pain and its implementation in the MET system. *Infor* 43:287–301
- Miller GA (1956) The magical number seven, plus or minus two: some limits in our capacity for processing information. *Psychol Rev* 63:81–97
- Pałkowski Ł, Błaszczyński J, Skrzypczak A, Błaszczak J, Kozakowska K, Wróblewska J, Kożusko S, Gospodarek E, Krysiński J, Słowiński R (2014a) Antimicrobial activity and SAR study of new Gemini imidazolium-based chlorides. *Chem Biol Drug Des* 83:278–288
- Pałkowski Ł, Krysiński J, Błaszczyński J, Słowiński R, Skrzypczak A, Błaszczak J, Gospodarek E, Wróblewska J (2014b) Application of rough set theory to prediction of antimicrobial activity of Bis-quaternary imidazolium chlorides. *Fund Inform* 132:1–16
- Pałkowski Ł, Błaszczyński J, Skrzypczak A, Błaszczak J, Nowaczyk A, Wróblewska J, Kożusko S, Gospodarek E, Słowiński R, Krysiński J (2015) Prediction of antifungal activity of gemini-imidazolium compounds. *Biomed Res Int* 2015:article ID 392326
- Pawlak Z (1982) Rough sets. *Int J Comput Inform Sci* 11: 341–356
- Pawlak Z (1991) Rough sets. Kluwer, Dordrecht
- Pawlak Z, Słowiński R (1994) Rough set approach to multi-attribute decision analysis. *Eur J Oper Res* 72: 443–459
- Pawlak Z, Słowiński K, Słowiński R (1986) Rough classification of patients after highly selective vagotomy for duodenal ulcer. *Int J Man Mach Stud* 24:413–433
- Polkowski L (2002) Rough sets. Physica-Verlag, Heidelberg
- Roberts F (1979) Measurement theory, with applications to decision making, utility and the social sciences. Addison-Wesley, Boston
- Rockafellar RT, Uryasev SP (2000) Optimization of conditional value-at-risk. *J Risk* 2:21–42
- Rossi L, Słowiński R, Susmaga R (1999) Rough set approach to evaluation of stormwater pollution. *Int J Environ Pollut* 12:232–250
- Roy B (1996) Multicriteria methodology for decision aiding. Kluwer Academic Publishers, Dordrecht
- Roy B (1999) Decision-aiding today: what should we expect. Chapter 1. In: Gal T, Stewart T, Hanne T (eds) Advances in multiple criteria decision making. Kluwer, Boston, pp 1.1–1.35
- Roy B, Bouyssou D (1993) Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris
- Salo A, Keisler J, Morton A (eds) (2011) Portfolio decision analysis: improved methods for resource allocation, vol 162. Springer Science & Business Media, Berlin
- Shoemaker PJH (1982) The expected utility model: its variants, purposes, evidence and limitations. *J Econ Lit* 20:529–562
- Słowiński R (1993) Rough set learning of preferential attitude in multi-criteria decision making. In: Komorowski J, Ras ZW (eds) Methodologies for intelligent systems, LNAI 689. Springer, Berlin, pp 642–651

- Słowiński R, Zopounidis C (1995) Application of the rough set approach to evaluation of bankruptcy risk. *Int J Intell Syst Account Financ Manag* 4:27–41
- Słowiński R, Zopounidis C, Dimitras AI (1997) Prediction of company acquisition in Greece by means of the rough set approach. *Eur J Oper Res* 100:1–15
- Słowiński R, Greco S, Matarazzo B (2002a) Rough set analysis of preference-ordered data. In: Alpigini JJ, Peters JF, Skowron A, Zhong N (eds) *Rough sets and current trends in computing*, LNAI 2475. Springer, Berlin, pp 44–59
- Słowiński R, Greco S, Matarazzo B (2002b) Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle. *Control Cybern* 31:1005–1035
- Słowiński R, Greco S, Matarazzo B (2007) Dominance-based rough set approach to reasoning about ordinal data. Keynote lecture. In: Kryszkiewicz M, Peters JF, Rybiński H, Skowron A (eds) *Rough sets and intelligent systems paradigms*. LNAI 4585. Springer, Berlin, pp 5–11
- Słowiński R, Greco S, Matarazzo B (2014) Rough set based decision support. Chapter 19. In: Burke EK, Kendall G (eds) *Search methodologies: introductory tutorials in optimization and decision support techniques*, 2nd edn. Springer, New York, pp 557–609
- Stefanowski J (1998) On rough set based approaches to induction of decision rules. In: Polkowski L, Skowron A (eds) *Rough sets in data mining and knowledge discovery*, vol 1. Physica, Heidelberg, pp 500–529
- Steuer RE, Choo E-U (1983) An interactive weighted Tchebycheff procedure for multiple objective programming. *Math Program* 26:326–344
- Susmaga R, Słowiński R, Greco S, Matarazzo B (2000) Generation of reducts and rules in multi-attribute and multi-criteria classification. *Control Cybern* 29(4): 969–988
- Szeląg M, Greco S, Słowiński R (2014) Variable consistency dominance-based rough set approach to preference learning in multicriteria ranking. *Inf Sci* 277: 525–552
- Tsoukias A, Vincke P (1995) A new axiomatic foundation of the partial comparability theory. *Theor Decis* 39: 79–114
- Wierzbicki AP (1980) The use of reference objectives in multiobjective optimization. In: Fandel G, Gal T (eds) *Multiple criteria decision making, theory and applications*. Springer, Berlin, pp 468–486
- Wilk S, Słowiński R, Michalowski W, Greco S (2005) Supporting triage of children with abdominal pain in the emergency room. *Eur J Oper Res* 160:696–709
- Zionts S, Wallenius J (1976) An interactive programming method for solving the multiple criteria problem. *Manag Sci* 22:652–663
- Zionts S, Wallenius J (1983) An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions. *Manag Sci* 29:519–523

Knowledge Engineering from Email Archives

Arvind Srinivasan¹ and Chien-Chung Chan²

¹SplitByte Inc., Los Gatos, CA, USA

²Department of Computer Science,
The University of Akron, Akron, OH, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Email Format History
Email System
Email Security and SPAM Prevention
Email Clients
Email Archiving
Compliance/Surveillance
eDiscovery
Archiving Modes
Archiving Tasks
Mail Parsing
Storing of Mails
User Association
Indexing and Querying
Implicit/Hidden Knowledge in Email Archives
Future Directions
Bibliography

Glossary

Simple Mail Transport Protocol (SMTP) is the standard for delivering Internet emails among mail servers.

Multipurpose Internet Mail Extensions (MIME) is a set of standard documents for email message format. It allows multiple objects of independent type to be mixed in an email message.

Email Spoofing is the creation of email messages with a forged sender address.

Technology-Assisted Review (TAR) also called predictive coding is the use of computer programs to spot documents that may be relevant to legal proceedings.

Electronic Discovery (eDiscovery) refers to discovery of electronically stored information (ESI) in legal proceedings such as litigation, government investigations, or Freedom of Information Act requests. Electronic discovery is subject to rules of civil procedure and agreed-upon processes, often involving review for privilege and relevance before data are turned over to the requesting party.

Legal Hold is a process that an organization uses to preserve all forms of relevant information when litigation is reasonably anticipated.

Definition of the Subject

Every enterprise employs electronic mail (email) as a primary communication tool between its employees, customers, partners, and vendors. Although more recent communication platforms are employed as well – including instant message, text message, Bloomberg, Reuters, blogs (Parlano), Slack, Hangout, Sametime, etc. – email usage has still increased significantly over the years. Additionally, nearly all collaboration activities within an enterprise leave trails within its email system. Even important content that typically appears in the form of a file such as legal contracts, invoices, sales conversion, and HR conversations is often embedded in emails due to collaboration and sharing of documents.

Email archives have become rich sources for data mining and data science studies. The scope of this entry is to review the process and practice of email archiving and data mining tasks that may be used to optimize the processes of business operation and management, to provide better customer service and support, to identify experts within an organization, to support regulatory compliances, to support eDiscovery processes for legal proceedings, and so on.

Introduction

Even though short message service (SMS) and instant messages (IM) are popular and widely used in social interactions, email remains the major form of communication in consumer and work-related communications. According to the Email Statistics Report released by the technology market research firm The Radicati Group, the number of worldwide email users will top 3.8 billion in 2018 and is expected to grow to over 4.2 billion by the end of 2022 (<https://www.radicati.com>). In 2018, the number of emails sent and received per day totals over 281 billion. This figure is expected to grow at an average annual rate of 3% over the next 4 years, reaching over 333 billion by the end of 2022. Email remains a key component of the online experience, as email accounts are required for any form of online activity ranging from signing up to social networking sites (e.g., Facebook and Twitter), accessing chat or instant messaging services, online shopping, or any other type of online activity (<https://www.radicati.com>).

Email archives are collections of email messages and attachments. Collectively, email messages embody not only the exchange of useful information and documents but also behavioral and communication patterns reflecting organizational and human social relations. They include explicit and hidden exchanges of information and knowledge in various formats such as text, images, audio, and video. They also include undesirable and harmful data such as spam, viruses, phishing, and fraudulent traps. The structure of email messages may be complex sequences or threads of messages involving a group of users from one or multiple organizations. Email messages have become important sources for legal proceedings such as litigation and government investigations. They can be used to derive useful knowledge to support business operations and management and customer service and support and discover areas of employee expertise. Email archives are rich and valuable data sources for knowledge engineering.

The process of email archiving is more than putting email messages and attachments into

storage. It is a complex process comprising differing levels of complexities. At the basic level, it is a process of acquire-store-retrieve: acquiring email messages from mail servers, storing messages to storage systems, and retrieving messages to meet business, regulatory, and legal requests. As the message formats become more complex and the email volumes increase to a large scale, there are many technical challenges and considerations for each step of the basic process: how to parse contents with multiple languages, which parts of an email message should be stored, how to store and index the archives to minimize retrieval and search latency, and how to deal with large volume of messages. At the advanced level, on top of search and retrieval operations, email archives are used to support data mining tasks to extract knowledge and meet organizational needs such as identifying customer sentiment, identifying employee morale, routing email questions to the respective expert, etc.

In the following section, we review the existing standards for formatting, delivering, and accessing email messages. We also explore popular standards related to email security. These standards are essential for processing email messages to build email archives. Next, we present the basic tasks of the email archiving process derived from the working knowledge of the field. Finally, we review the application of data mining tasks to extract implicit and hidden knowledge from email archives.

Email Format History

Email originated as a communication format for people to share simple text, but it soon expanded to sharing of attachments through one of many ad hoc methods such as uuencode (Unix-to-Unix encoding), written by Mary Ann Horton at UC Berkeley in 1980 (<http://www.tuhs.org/cgi-bin/utree.pl?file=4BSD/usr/man/cat1/uuencode.1c>). Technically, email messages are viewed as having an *envelope* and *content*. The envelope contains whatever information is needed to accomplish transmission and delivery, and the Simple Mail Transfer Protocol (SMTP) was proposed in the

Request for Comments (RFC) document RFC 821 in 1982 (<https://www.rfc-editor.org/rfc/rfc821.txt>). The content of an email composes the object to be delivered to the recipient, and the first standard for the format of email messages was proposed in RFC 822 in 1982 (<https://www.rfc-editor.org/rfc/rfc822.txt>). A more recent revision is given in RFC2822 (<https://www.rfc-editor.org/rfc/rfc2822.txt>).

In RFC 822, the format is specified for text messages where a message consists of header fields and, optionally, a body. A header field consists of a field name and a field body separated by the colon “:” symbol. Field name must be printable ASCII characters, and field body can be any ASCII characters except CR (carriage return) or LF (line feed). The only required header fields are the origination date field and the originator address field. The body is simply a sequence of lines containing ASCII characters, and it is separated from the headers by a null line, namely, a line with nothing preceding the CRLF.

As the Internet has continued to evolve, advancements in HTML and rich content creation have also spread into the email world. The Multipurpose Internet Mail Extensions (MIME) standard for email format was introduced in 1996, RFC 2045 (<https://www.rfc-editor.org/rfc/rfc2045.txt>), with other five linked RFCs: RFC 2046, RFC 2047, RFC 2049, RFC 4288, and RFC 4289. All the RFCs are available from the website <https://www.rfc-editor.org>. MIME allows email message body to contain multiple objects of independent type in a single message called multipart, non-English languages, multiple fonts, and multimedia objects such as images, audio, and video. Non-ASCII characters can be used in header fields and message body.

The MIME standard along with Base64 and Quoted-Printable encodings (<https://www.rfc-editor.org/rfc/rfc2045.txt>) for content enabled a standardized method of exchanging text and attachments. Early email clients, i.e., software used to access and send emails, lacked the ability to render HTML. The MIME standard creates the multipart/alternative standard to allow a text component and an HTML component, giving the email client an option to render the best content

it can handle. Even with HTML having become a standard for almost all email clients, the practice of keeping plain text along with HTML has become a standard practice to this date. Typical email textual content can be described as unstructured data. However, it may include structured parts such as greetings, signatures, disclaimers, etc. Attachments containing structured files such as spreadsheets may have unstructured notes of text, voice, etc.

Initially, email headers were limited to a handful of standard headers such as From, To, Subject, and Date. As email has evolved, the header scope has expanded to several standard headers, including Cc, Bcc, Sender, In-Reply-To, etc. Routing information was also added to the header to display the origin of an email. The RFC standard describes several standard headers and has provided the capability to add any number of ad hoc headers. The general practice is to start the non-standard header as “x-”. Some header field values are structured data, such as Date, Sender, and Recipient addresses. Other header field values are getting longer and longer, and some of them has become unstructured values.

The address format itself evolved from simple username to the current Internet mail address which has the format `user@xyz.com` with the latter part following the “@” symbol known as the domain. Many enterprise mail systems have their own internal user names, as in Microsoft Exchange server using “LegacyExchangeDN” address format, and Lotus Notes has its own address format. However, all systems convert to the Internet mail address scheme when they exchange emails using the MIME standard.

Replies and forwards were once very simple, with some portion of text along with some headers included along with the reply/forward. This has evolved (due to the RFC standards) to where the whole email can be and is treated as an attachment. The nested nature of MIME enables creation of very complex email structures. This constant reply/forward creates message threads that can present a set of technological challenges.

The following examples demonstrate particular structural complexities of email messages using the MIME standard. Under the RFC 822 format, email messages have simple structures but

```
From: John Doe <JohnDoe@gmail.com>  
Subject:  
Date: Wed, 5 Sep 2018 15:35:25 -0400  
To: John Doe <JohnDoe@gmail.com>
```

Knowledge Engineering from Email Archives, Fig. 1 The simplest email with an empty body, and it has sender's address and Date

```
From: John Doe <JohnDoe@gmail.com>  
Subject: Greeting  
Date: Wed, 5 Sep 2018 15:45:46 -0400  
To: John Doe <JohnDoe@gmail.com>
```

Hello,

Have a good day!

Knowledge Engineering from Email Archives, Fig. 2 Typical email with a plain text body, sender's address, Subject, and Date

are limited to ASCII printable characters only. In Fig. 1, the simplest email has an empty body, and only the sender's address and Date headers are required. In Fig. 2, Subject and plain text body are added to the email message. Figure 3 highlights a message with the MIME format shown. The complexity of the email message increases dramatically where the sender's address contains some foreign language characters, and the body contains two parts: one in plain text with foreign characters and the other in HTML format. The Quoted-Printable encoding is used for both parts as Content-Transfer-Encoding, and a boundary delimiter string is used to separate the different parts in the message body.

It is common to forward emails to other recipients. The "References:" header is used to store the Message-Id of the message being forwarded as shown in Fig. 4, where the Subject has the value "Fwd: Greeting," and the body of the forwarded email is shown after the "Begin forwarded message:" header.

Email System

Email transport works in various ways. The standard method of delivery occurs by using the Simple Mail Transfer Protocol (SMTP), and a recent standard is specified in RFC 5321 (<https://www.rfc-editor.org/rfc/rfc5321.txt>).

The steps to send and receive mail are:

- In Fig. 5, the sender uses his email client to create a message and click the send command, which transmits the mail to the local mail server.
- The local mail server will use the SMTP protocol to deliver to the destination. The local mail server has a software component called mail transfer agent (MTA), which implements both the client (sending) and server (receiving) portions of the protocol. The MTA uses the recipient SMTP email address as well as the domain portion of the SMTP address, and it finds the official receiving SMTP server for that domain by looking up the MX records, which are defined using the Domain Name System (DNS) infrastructure. The MTA then establishes a TCP connection and conducts the SMTP session protocol to deliver the mail.
- The receiving SMTP server will now either store the message for the user or forward it to the appropriate mail server which hosts the addressed recipient's mailbox.

Email Security and SPAM Prevention

The Internet mail exchange carries many fundamental issues that one sees with the conventional snail mails. Many emails have malicious

```

Delivered-To: JohnDoe@gmail.com
.....
From: "?=utf-8?B?77u/77u/5b+r5LiJ57ay6Lev5ZWG5Z+O?="<service@edm.tk3c.com.tw>
To: <JohnDoe@gmail.com>
MIME-Version: 1.0
Content-Type: multipart/alternative; boundary=0005868a568-4e54-4f93-ac58-62ab8fd60baf

--0005868a568-4e54-4f93-ac58-62ab8fd60baf
Content-Type: text/plain; charset=utf-8
Content-Transfer-Encoding: quoted-printable

<http://edm.tk3c.com.tw/HL/8fd572/ce0b87b/5236fcf/a1ac1/a1877/a18d4/1/1311/=
800.htm>
=B5o=B0e=A4=E9=A1G2018=A6~9=A4=EB5=A4=E9
<http://edm.tk3c.com.tw/HL/8fd57a/ce0b87b/5236fcf/a1ac1/a1877/a18d4/1/1311/=
800.htm>

--0005868a568-4e54-4f93-ac58-62ab8fd60baf
Content-Type: text/html; charset=utf-8
Content-Transfer-Encoding: quoted-printable

=EF=BB=BF<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "ht=
tp://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns=3D"http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv=3D"Content-Type" content=3D"text/html; charset=3DUTF-8" />
<title>=E7=87=A6=E5=9D=A4=E7=B6=B2=E8=B7=AF=E6=97=97=E8=89=A6=E5=BA=97EDM<
/=
title>
</head>
<style type=3D"text/css">
img{border:none;}
body {margin-top: 0px;background-color: #FFF;}
</style>

<body>

<table width=3D"954" border=3D"0" align=3D"center" cellpadding=3D"0" cellspacing=
3D"0">
<tr>
.....

```

Knowledge Engineering from Email Archives, Fig. 3 Email with MIME format allowing foreign language and multipart of bodies

attachments or HTML body with JavaScript that can be source of viruses and other malware. Therefore, it is very common for mail systems to scan email messages for viruses, spam, and forms of fraud such as phishing (note here that mail spam refers to unsolicited messages sent in bulk, and phishing refers to the use of fraudulent messages to obtain a recipient's sensitive information such as usernames, passwords, or credit card numbers by disguising as a trustworthy sender).

Another common threat is email spoofing, which describes the creation of email messages using a forged sender address. It is common for spam and phishing emails to use spoofing to mislead the recipient concerning the origin of messages. Although email spoofing is effective in forging the email address, the IP address of the computer sending the mail can generally be identified from the "Received:" email header field. In many cases, it is likely to be an innocent third

```

From: John Doe <JohnDoe@gmail.com>
Content-Type: multipart/alternative;
boundary="Apple-Mail= _B330B885-9355-4748-B95D-0D421A8DF778"
Mime-Version: 1.0 (Mac OS X Mail 11.5 \((3445.9.1)\))
Subject: Fwd: Greeting
X-Universally-Unique-Identifier: BE9E6AA7-C699-4A10-A7B1-9B733F12BC2A
Message-Id: <2BB70B80-52C5-4BAD-BED5-6FF80745DDD6@gmail.com>
References: <3877961D-91C5-44F5-BAE1-40E15308D04A@gmail.com>
To: John Doe <JohnDoe@gmail.com>
Date: Wed, 5 Sep 2018 16:18:57 -0400
--Apple-Mail= _B330B885-9355-4748-B95D-0D421A8DF778
Content-Transfer-Encoding: 7bit
Content-Type: text/plain;
charset=us-ascii
> Begin forwarded message:
>e
> From: John Doe <JohnDoe@gmail.com>
> Subject: Greeting
> Date: September 5, 2018 at 3:45:46 PM EDT
> To: John Doe <JohnDoe@gmail.com>
>
> Hello,
>
> Have a good day!
--Apple-Mail= _B330B885-9355-4748-B95D-0D421A8DF778
Content-Transfer-Encoding: quoted-printable
Content-Type: text/html;
charset=us-ascii
<html><head><meta http-equiv=3D"Content-Type" content=3D"text/html; =
charset=3Dus-ascii"></head><body style=3D"word-wrap: break-word; =
-webkit-nspace-mode: space; line-break: after-white-space;" class=3D""><br =
class=3D""><div><br class=3D""><blockquote type=3D"cite" class=3D""><div =
.....
class=3D""></span></div><br class=3D""><div class=3D""><div =
class=3D"">Hello,<br class=3D""><br class=3D"">Have a good day!<br =
class=3D""></div></div></blockquote></div><br class=3D""></body></html>=
--Apple-Mail= _B330B885-9355-4748-B95D-0D421A8DF778--

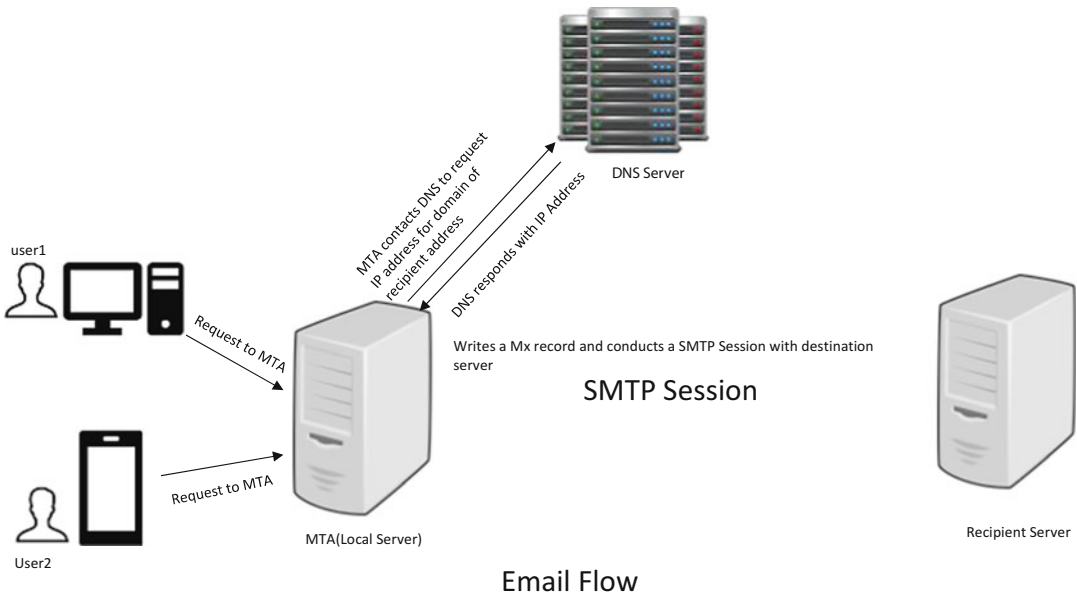
```

Knowledge Engineering from Email Archives, Fig. 4 This email forwards another email

party infected by malware that is sending the email without the owner's knowledge.

Detecting spam is one of the earliest machine learning problems, and initially this was solved using the simple Naïve Bayes algorithm (Sahami et al. 1998). However, spammers started using more sophisticated techniques to defeat spam detectors, thus creating a cat and mouse game that has sustained for several years. Numerous research articles have appeared on this subject, and many commercial and open-source programs have been developed and actively used to combat spam (<https://spamassassin.apache.org>; Abu-Nimeh et al. 2007; Graham 2003; Pantel and Lin 1998; Yerezunis 2003).

Additionally, several rules have been introduced to prevent spammers. Many mail servers now require the sending mail server to have a reverse DNS entry. Additionally, consistency checks are performed to verify if the "From:" address and the "Return-Path:" headers have a matching or consistent address. Previously, domains associated with spam were collected and added to databases that many mail servers would blacklist. While this reduced spam, legitimate marketing campaigns would frequently be incorrectly tagged as spam. As a result, new protocols and best practices were eventually developed by legitimate email marketers. For example,



Knowledge Engineering from Email Archives, Fig. 5 The flow of transporting an email

many provided a method to unsubscribe to automated messages.

Email spoofing is closely related to spamming, and there are some widely used protocols for dealing with spoofing. To effectively prevent forged email from being delivered, the sending domains, their mail servers, and the receiving system all need to be configured correctly for these higher standards of authentication. They are described in the following.

Sender Policy Framework (SPF) is given in RFC 7208 (<https://www.rfc-editor.org/rfc/rfc7208.txt>) and is designed to detect and block email spoofing by allowing receiving mail servers to verify that the IP address of incoming mails from a domain is authorized by that domain's administrators. The list of authorized sending hosts and IP addresses for a domain is published in the DNS records for that domain in the form of a specially formatted TXT record. Email spam and phishing often use forged "From:" addresses and domains, so publishing and checking SPF records can be considered one of the most reliable and simple to use anti-spam techniques.

DomainKeys Identified Mail (DKIM) is specified in RFC 6376 (<https://www.rfc-editor.org/rfc/rfc6376.txt>), and it is an email authentication

method designed to detect email spoofing. It is intended to prevent forged sender addresses in emails. It allows the receiver to check that an email claimed to have come from a specific domain was indeed authorized by the owner of that domain. DKIM allows hosts of a specific domain to attach digital signatures to email messages sent from the domain. Verification is carried out using the signer's public key published in the DNS. A valid signature guarantees that some parts of the email (possibly including attachments) have not been modified since the signature was affixed by mail server host systems rather than the message's authors and recipients.

Domain-based Message Authentication, Reporting, and Conformance (DMARC) is given in RFC 7489 (<https://www.rfc-editor.org/rfc/rfc7489.txt>). It is intended to combat certain techniques often used in phishing and email spam, such as emails with forged sender addresses that appear to originate from legitimate organizations. Specified in RFC 7489, DMARC counters the illegitimate usage of the exact domain name in the "From:" field of email message headers. DMARC doesn't directly address whether or not an email is spam or otherwise fraudulent. Instead, DMARC requires that a message not only pass

DKIM or SPF validation but that it also pass alignment. Under DMARC, a message can fail even if it passes SPF or DKIM, but fails alignment. DMARC operates by checking that the domain in the message's "From:" field (also called "5322.From") is aligned with other authenticated domain names. If either SPF or DKIM alignment checks pass, then the DMARC alignment test passes.

The results of applying security filtering are recorded in email headers. As shown in Fig. 6, complexity of email headers increases dramatically where a "Return-Path:" header is used to store the sequence of MTAs used to relay the message. "Authentication-Results:" from [Google.com](https://www.google.com) is recorded as the example was created using Gmail, and a "DKIM-Signature:" was affixed by the sender's mail server.

Email Clients

Email clients have evolved too. Early clients simply polled the mail server and downloaded the messages locally before deleting them from servers using a protocol POP3 (Post Office Protocol Version 3) (<https://www.rfc-editor.org/rfc/rfc1225.txt>). The mail server is simply treated as temporary store for the user, and the mail client is expected to store and provide functions such as foldering. With email becoming more central to work activity, these mails were needed on multiple clients. Organization started providing larger storage to individual mailboxes resulting in new requirements such as folders and mailbox synchronization. The more complex IMAP (Internet Message Access Protocol) (<https://www.rfc-editor.org/rfc/rfc3501.txt>) became the open

```
Delivered-To: JohnDoe@gmail.com
Received: by 2002:a9f:260e:0:0:0:0 with SMTP id 14-v6csp5102228uag;
  Wed, 5 Sep 2018 07:58:06 -0700 (PDT)
X-Google-Smtp-Source:
ANB0VdY6dl5tQYWzm025rVa8L4nDZJcpMUTKjVmxbpPf/DK5a7IC1aZVpyuV5EjB3pNxBe6pzUs
H
X-Received: by 2002:aca:487:: with SMTP id 129-v6mr31699002oie.275.1536159486031;
  Wed, 05 Sep 2018 07:58:06 -0700 (PDT)
Return-Path: <Return.EIDce0b87b.Job@edm.tk3c.com.tw>
Received: from mx3.tkec.com.tw ([122.147.9.85])
  by mx.google.com with ESMTPS id 8-v6si1252942oix.218.2018.09.05.07.58.04
  for <JohnDoe@gmail.com>
  (version=TLS1_2 cipher=ECDHE-RSA-CHACHA20-POLY1305 bits=256/256);
  Wed, 05 Sep 2018 07:58:06 -0700 (PDT)
Received-SPF: pass (google.com: domain of return.eidce0b87b.job@edm.tk3c.com.tw designates
122.147.9.85 as permitted sender) client-ip=122.147.9.85;
Authentication-Results: mx.google.com;
  dkim=pass header.i=@edm.tk3c.com.tw header.s=s1024 header.b=e4e695Yi;
  spf=pass (google.com: domain of return.eidce0b87b.job@edm.tk3c.com.tw designates
122.147.9.85 as permitted sender) smtp.mailfrom=Return.EIDce0b87b.Job@edm.tk3c.com.tw;
  dmarc=pass (p=NONE sp=NONE dis=NONE) header.from=edm.tk3c.com.tw
DKIM-Signature: v=1; a=rsa-sha1; c=relaxed/relaxed; q=dns/txt;
  d=edm.tk3c.com.tw; s=s1024;
  h=message-id:from:mime-version;
  .....
Subject: =?utf-
8?B?77u/77u/44CQ5YCS5pW4MuWkqSDmjlflrpppUGhvbmuUgODjmiipjgJHonqLluZU35oqYdXDlho
3nj77mipjmnIDpq5gx5Y2D77yM5ZyL6Zqb54mM6bq15YyF5qmf54++5oqYMuWNg+iOiOWDUEWAW
Dk5OeWFg+mkkOWFt+e1hO+8gQ==?=
From: "=?utf-8?B?77u/77u/5b+r5LiJ57ay6Lev5ZWG5Z+O?="<service@edm.tk3c.com.tw>
To: <JohnDoe@gmail.com>
MIME-Version: 1.0
Content-Type: multipart/alternative; boundary=0005868a568-4e54-4f93-ac58-62ab8fd60baf
.....
```

Knowledge Engineering from Email Archives, Fig. 6 Email headers showing security filtering results

Internet standard for email clients to synchronize with the server at a folder level.

Enterprise class mail servers such as MS Exchange and Lotus Notes had their proprietary RPC (Remote Procedure Call)-based protocols to enable richer client needs. With the development of web, many free web mail services such as RocketMail, Hotmail, Yahoo Mail, Gmail, etc. evolved. While emails were exchanged via standard SMTP between mail servers, email client interactions including sending, receiving, and organizing mails were all built with vendor-specific methods. Soon enterprise mail servers like Exchange and Lotus Notes also provided web interfaces. Email-specific client devices such as Blackberry became the most important client due to user demand for instantaneous access to email, and the capability of such devices provided users to compose new emails. With the advent of smartphones and their use of standard email clients, this has become the most common method to access email. However, thick clients such as Outlook and Lotus Notes are still used significantly along with the web interface. In addition, email providers have their own smart phone clients.

Coinciding with these evolutions was a trend toward more and more emails being kept on mail servers, as well as the continued organization of emails by users into folders, views, and tags. Additionally, content within the mail servers now started to contain other related information such as appointments, contact, tasks, and notes.

Email Archiving

Email data has been steadily building in volume within email servers. Users have found it useful to keep historical emails for future reference and therefore began organizing their mailboxes. Every organization began to see the need for large mailboxes. In addition, corporations started routinely archive the mailboxes of former employees to provide business continuity. The collective information within the various mailboxes provides a view into the behavior of individuals, departments, and the organization as a

whole. More specifically, enterprises can leverage email data to support regulatory compliance and potential legal proceedings.

Compliance/Surveillance

The very first need for reviewing email at a corporate level originated from the requirement to perform surveillance on the communication of a certain set of key employees. Impacting all broker dealer communication with customers and other key stake holders, SEC Rule 17a-4 requires financial institutions to monitor content for potential compliance violations and issues relating to customer interactions that could result in excessive/aggressive sales techniques, customer complaints, and inappropriate claims of financial instrument performance (<https://www.sec.gov/rules/final/34-44992.htm>).

eDiscovery

In the United States, electronic discovery was the subject of amendments to the Federal Rules of Civil Procedure (FRCP), effective December 1, 2006, as amended on December 1, 2015 (<http://www.uscourts.gov/rules-policies/current-rules-practice-procedure/federal-rules-civil-procedure>).

Following the landmark case *Zubulake v. UBS Warburg*, the Department of Justice introduced a new amendment relating to electronic documents. The amendment essentially made evidence found in electronic documents explicitly admissible and equal to other physical evidence in the court of law. As a result, it places burden on the enterprise to collect, preserve, and produce electronic information and treat it as an evidence.

Emails are perfect information for the purposes of surveillance, investigation, and eDiscovery, because emails provide the rare records of the following types of information:

- Sender: Indisputable in corporate environment as they use their login credential to send emails.

- **Date/Timestamp:** The Sent Date header along with dates in the routing headers clearly pin the time without any doubt.
- **To, Cc, Bcc:** These headers when available will clearly indicate all the parties involved.
- **Replies, Forwards, and Conversation Threads:** These headers provide a body of evidence leading up to the event of interest. It is very difficult to separately document all the emails while maintaining consistency. This is especially true when the other party has copies of some of relevant conversation, thereby providing a valid consistency check.

Archiving Modes

With the requirements of compliance, eDiscovery, and recordkeeping, the volume of emails within the mail server skyrocketed. It becomes very challenging to keep the mail server operational with desired performance while performing functions such as making backups and having a disaster solution in place. In addition, with the introductions of eDiscovery, organizations found that they could not delete emails related to any potential lawsuit or dispute. To combat this, organizations began to deploy archive solutions in which historical and production emails are maintained in one or many archive systems. Depending on organizations and their needs, archiving of emails can occur in one of the following ways:

- Acquire all messages as they occur – (also known as journaling). Mail servers manage to provide a stream of emails while they are transferred from one mailbox to another. Proactive capture needs to keep up with email volumes as they're created (previously sent emails cannot be obtained through this method).
- Acquire messages directly from mail server by polling – (also known as mailbox crawling). It is a slow and tedious process. It places burden on mail servers, and efficiency is dependent on the synchronization capabilities.
- Manually collect the client files stored as PST, NSF or loose MSG, MIME files.

The capture techniques used to collect all emails from a mail server are challenging, and one can run into many issues with both mail servers' scaling capabilities and capturing clients' scaling techniques. Once the data is acquired, there are options to store them in either native or normalized form (such as MIME). Depending on the use case, system will choose to store in one or both forms.

Archiving Tasks

There are many basic tasks in the email archiving process, and these include parsing, storing, user association, indexing, and querying. The tasks are further described in the following.

Mail Parsing

The MIME standard provides great extensions and flexibilities for email formats. Parsing of mails starts with extracting key fields such as From, To, Cc, Bcc, Date, Subject, other headers, message body, and attachments. Each of these fields are themselves compound fields. The RFC's specifications do provide valuable insight into understanding the various fields. Due to the evolution of email, headers are expected to be 7-bit. Therefore, parsing headers of non-Latin emails such as Japanese and Chinese add complexity to the parser requirement. The message date is in a header field called Date; however, it reflects the date created by the email client and therefore may be untrustworthy, especially with external emails where users often use a future date in order to keep emails on the top of the inbox. In such situations, other dates can be inferred from the Received header. There is only a unique identifier that is set on Message-ID header field. Though the majority of commercial email clients do generate a globally unique Message-ID, many real-life emails may not have the Message-ID header itself, and in certain situations, they may not be unique either. Other header fields such as In-Reply-To may refer to the Message-ID from a related earlier email. These header fields can be used to form email threads.

Parsing of the body and attachment, though complex, in most cases is straightforward and requires knowledge of encoding standard Base64 and Quoted-Printable encoding. In addition, a single body can become extremely long because it may contain content from previous emails in situations when it is a Reply or Forward. The message body itself can be plain text or HTML. This requires HTML to Text conversion.

Email clients may attach original mails such as a nested mail, where the attachment itself is a valid MIME message itself. There are a variety of attachments; some are given in the following:

- **Embedded or Reference Attachment:** This usually refers to attachments (usually images) that are directly referenced in the body and therefore operate more so as part of the body rather than attachment.
- **Inline Attachment:** This is a regular attachment with a hint to the rendering email client to render the attachment along with the body.
- **Normal Attachment:** This is a standard attachment that is expected to be downloaded. Usually the file name and the content type are specified in the attachment header field.
- **Nested Mail:** This is a complete mail that usually has been forwarded or replied to.

Many attachment documents are binary in nature such as images, MS Office documents, PDF, audio, video, zip, tar, etc. Many commercial and open-source software exist to extract text from standard Office and PDF and other word processing documents. With regard to extracting texts from images and PDF, it requires OCR and handwriting recognition software. Similarly, extracting texts from audio and video is a challenging and time-consuming task. Moreover, extracting files from container files such as zip, tar, rar, etc. makes the task of extraction of text from attachments a very a complex and deeply nested task.

Storing of Mails

Storing of emails in both MIME form and native format at large volumes is challenging. Many

applications require the stored emails to be tamperproof. Archive solutions typically compute digest of the content and associate it with the content or even better encrypt the messages to keep the messages from being tampered. One of the challenges of processing emails is that they're often very small (a few KB) with occasional large mails going to hundreds of MB. The quantity of emails for a reasonably large enterprise can reach the billions, and dealing with a very large number of small emails will cause the email processing system to be burdened with heavy overhead.

Usually, emails are de-duplicated to reduce the storage requirement. There are two sources of duplication. The message itself is a duplicate (the same email fetched from multiple mailboxes, or journal capture at different path may produce duplicate emails). The second source of duplicates is when the same attachment is added to different emails. To process deduplication, hashes are constructed for attachment and message. For message, some normalization is done prior to hash construction. The constructed hashes are checked against previously stored hashes, and if one is found, we add a reference to the existing object; otherwise a new object is created, and we store the hashes for future deduplication.

User Association

One of the fundamental tasks of email processing is to associate both internal and external users to mails. The mails have Sender, To, Cc, and Bcc headers that identify the users. However, extracting email address alone is not always enough to associate mails to an actual user. Users can have multiple addresses, organizations can have multiple domain names, and users' email addresses can change due to name changes. Associating users to emails in real time is even more challenging as new users' email addresses are not readily known to the processing system. The lookup table for email addresses can become fairly large, and keeping them in synchronization with reality is challenging. Many data mining and application tasks need to have accurate user and organization association with emails.

Indexing and Querying

The task of indexing an organization's emails and their content is a very challenging one. Indexing needs to occur in parallel and potentially to be stored into multiple partitions. Parsing of emails into fields, extracting text from body and attachments, and enriching the raw index with derived fields such as user association, organization association, and other tags will make the indexed information extremely useful for downstream querying.

Building a corporate level email archive brings many security challenges as it is required to discern management and customer emails, and not all emails should be readily searchable by anyone and everyone. Strict access control along with adding restrictions based on the search user is essential.

There are many specialized search needs depending on the purpose of an email archive. However, most of the queries are typical searches as found in the following:

- Simple metadata searches by users within a time frame
- Simple Boolean keyword searches along with other users and date constraints
- More complex searches such as wild card search, fuzzy search, phrase search, and proximity search
- Highly complex searches that include span query and wild card proximity search (e.g., the keywords *sell** and *stock** found within five words of one another)

Implicit/Hidden Knowledge in Email Archives

Data mining from large data sets has become a common practice in many enterprises. Email archives are rich data sources for data mining tasks such as association, classification, clustering, sentiment analysis, social network analysis, and speech acts detection (Carvalho 2011; Cohen et al. 2004; Ducheneaut and Watts 2005; Jiawei and Kamber 2011; Liu 2006; Scott and Carrington

2011; Tan et al. 2006; Tang et al. 2014). Knowledge extracted from the message content and attachments can provide valuable insights for enterprises to enhance their workflows.

For business operation/management, association rules and social network analysis can be used to generate an expertise power graph showing who works on what, and an organization power graph can be mined showing who knows who, key influencers, and working relationships of leaders/subordinates (Rowe et al. 2007; Schwartz and Wood 1993; Trusov et al. 2010; Tyler et al. 2003). With topic modeling and association rule mining, automatic expertise discovery can be accomplished (Blei 2012; Campell et al. 2003; Cao et al. 2006; Fang and Zhai 2007; Mimno and McCallum 2007). Detection of employee energy level, morale, and sentiments can be used to boost productivity (Hutto and Gilbert 2014; Liu 2010). Speech acts detection from email messages can be used for coordinating activities, meetings, and collaborations (Carvalho 2011). eDiscovery tools can be used for monitoring regulatory compliance and supporting litigation (Casey 2009; Cormack and Grossman 2014; Oard and Webber 2013; Roitblat et al. 2010; Vinjumur et al. 2016). Detection of information leaks can minimize an enterprise's potential loss and provide improvement on operational workflow (Carvalho 2011).

Customer services and supports can benefit from real-time detection of customer facing problems. Expertise power graph can be used to route customer requests and questions to the right expert (Balog et al. 2006; Campell et al. 2003). Sentiment analysis can be used to detect customer sentiment and enable the success of sales departments.

Email segmentation is used to identify different sections consisting of a message body such as greetings, main body, signature, and disclaimer. It can be used to avoid storing duplicated information. Machine learning algorithms have been applied for email segmentation (Berger et al. 1996; Carvalho and Cohen 2002; Lampert et al. 2009). This is a common practice to organize emails by threads or conversations. Detection of email thread chains is a useful process to support email search, summarization, classification, and

Knowledge Engineering from Email Archives, Table 1 Publication references for data mining tasks

	Association	Classification	Clustering	Sentiment analysis	Social network analysis	Speech acts detection	Topic modeling	eDiscovery	References
Business operation/management									
Expertise power graph	X				X		X		(Balog et al. 2006; Blei 2012; Campell et al. 2003; Cao et al. 2006; Fang and Zhai 2007; Li et al. 2004; Mimmo and McCallum 2007)
Organization power graph (leaders/subordinates)	X				X				(Li et al. 2004; Rowe et al. 2007; Schwartz and Wood 1993; Tyler et al. 2003)
Key influencers	X				X				(Cao et al. 2006; Fang and Zhai 2007; McCallum et al. 2007; Mimmo and McCallum 2007; Trusov et al. 2010)
Automatic expertise discovery			X				X		(Balog et al. 2006; Blei 2012; Campell et al. 2003; Cao et al. 2006; Fang and Zhai 2007)
Measuring employee energy level, morale/sentiments		X		X					(Cha and Cho 2012; Liu 2006; Otte and Rousseau 2002; Scott and Carrington 2011)
Activity/task coordination: meetings, collaborations,		X				X			(Carvalho 2011; Carvalho and Cohen 2002; Cohen et al. 2004)
Regulatory compliance, litigation		X						X	(Casey 2009, Cormack and Grossman 2014; Oard and Webber 2013; Roitblat et al. 2010; Vinjumar et al. 2016)
Information leaks		X				X	X		(Blei 2012; Carvalho 2011)
Customer services/supports									

(continued)

Knowledge Engineering from Email Archives, Table 1 (continued)

	Association	Classification	Clustering	Sentiment analysis	Social network analysis	Speech acts detection	Topic modeling	eDiscovery	References
Real-time detection of customer facing problems		X				X	X		(Blei 2012; Carvalho 2011)
Routing emails questions to the right expert	X					X	X		(Bekkerman 2018; Blei 2012; Carvalho 2011)
Customer sentiment		X		X					(Hutto and Gilbert 2014; Liu 2010, 2012; Pang and Lee 2008; Pang et al. 2002; Turney 2002; Turney and Littman 2003)
Sales team success	X			X					(Liu 2010; Pang and Lee 2008; Turney 2002; Turney and Littman 2003)
Email processing									
Single instance			X						(Arasu et al. 2009)
Email segmentation		X							(Berger et al. 1996; Carvalho 2011; Carvalho and Cohen 2002; Lampert et al. 2009)
Signature extraction		X							(Carvalho and Cohen 2002)
Email thread chain			X						(Ailon et al. 2013; Di Castro et al. 2018; Yeh and Hamly 2006)
Categorization		X				X	X		(Bekkerman 2018; Blei 2012; Carvalho 2011)
Similar documents			X				X		(Blei 2012; Radev et al. 2004; Salton et al. 1975)

visualization (Ailon et al. 2013; Di Castro et al. 2018; Yeh and Harnly 2006). Similar document detection can be used to reduce storage space (Arasu et al. 2009). Topic modeling can be used to categorize emails (Bekkerman 2018; Blei 2012).

The following table summarizes what data mining tasks can be applied to extract what types of knowledge from email archives (Table 1).

Future Directions

In the past few years, both private organizations and government agencies have built very large email archives primarily for eDiscovery, compliance, and recordkeeping. A variety of analytics dashboards can be built by leveraging data within an archive. Examples include:

- Identifying key people within the organization.
- Identifying areas of expertise as well as business continuity issues.
- Understanding employee sentiments by visualizing and analyzing email activity and response time during work day and off time. This can enable the management of employee flight risk.
- Reviewing sales activity via emails to understand sales success and sales failure.
- Identifying personally identifiable information (PII) of customers and employees and associating PII with customers and employees. This will help in satisfying requirements of the General Data Protection Regulation (GDPR).

The above examples give a tiny peak into potential analytics and dashboard applications. Many organizations are leveraging these kinds of information to gain competitive advantages over competitors. From a research perspective, all the techniques introduced in this article will need to evolve and scale to the ever-increasing volume of content. Also, many of the systems will move into the cloud, and the existing technology will need to evolve to work with the new cloud offerings.

Bibliography

- Abu-Nimeh S, Nappa D, Wang X, Nair S (2007) A comparison of machine learning techniques for phishing detection. In: Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit (eCrime'07). ACM, New York, pp 60–69
- Ailon N, Karnin ZS, Liberty E, Maarek Y (2013) Threading machine generated email. In: Proceedings of the sixth ACM international conference on web search and data mining (WSDM'13). ACM, New York, pp 405–414
- Arasu A, Ré C, Suciu D (2009) Large-scale deduplication with constraints using dedupalog. In: 2009 IEEE 25th international conference on data engineering, Shanghai, pp 952–963
- Balog K, Azzopardi L, de Rijke M (2006) Formal models for expert finding in enterprise corpora. In: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR'06). ACM, New York, pp 43–50
- Bekkerman R (2018) Automatic categorization of email into folders: benchmark experiments on Enron and SRI corpora. Computer Science Department Faculty Publication Series, 218
- Berger A, Della Pietra VJ, Della Pietra SA (1996) A maximum entropy approach to natural language processing. *Comput Linguist* 22(1):39–71
- Blei DM (2012) Probabilistic topic models. *Commun ACM* 55(4):77–84
- Campell CS, Maglio PP, Cozzi A, Dom B (2003) Expertise identification using email communications. In: CIKM 2003. ACM, New York, pp 528–531
- Cao Y, Liu J, Bao S, Li H (2006) Research on expert search at enterprise track of trec2005. In: Proceedings of TREC-05
- Carvalho VR (2011) Modeling intention in email: speech acts, information leaks and recommendation models. Springer, Berlin/Heidelberg
- Carvalho VR, Cohen WW (2002) Learning to extract signature and reply lines from email. CEAS 2004, first conference on email and anti-spam, Mountain View, 30–31 July 2004
- Carvalho VR, Cohen WW (2005) On the collective classification of email speech acts. In: Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR'05). ACM, New York, pp 345–352
- Casey E (2009) Handbook of digital forensics and investigation. Academic, Orlando
- Cha Y, Cho J (2012) Social-network analysis using topic models. In: Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval (SIGIR'12). ACM, New York, pp 565–574
- Cohen WW, Carvalho VR, Mitchell TM (2004) Learning to classify email into speech acts. In: Proceedings of the EMNLP, Barcelona

- Cormack GV, Grossman MR (2014) Evaluation of machine-learning protocols for technology-assisted review in electronic discovery. In: Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval (SIGIR'14). ACM, New York, pp 153–162
- Di Castro D, Gamzu I, Grabovitch-Zuyev I, Lewin-Eytan-L, Pundir A, Sahoo NR, Viderman M (2018) Automated extractions for machine generated mail. In: Companion proceedings of the web conference 2018 (WWW'18), Republic and Canton of Geneva, pp 655–662
- Ducheneaut N, Watts LA (2005) In search of coherence: a review of e-mail research. *Hum-Comput Interact* 20(1):11–48
- Fang H, Zhai CX (2007) Probabilistic models for expert finding. In: Amati G, Carpineto C, Romano G (eds) Proceedings of the 29th European conference on IR research (ECIR'07). Springer, Berlin/Heidelberg, pp 418–430
- Gomez JC, Boiy E, Moens MF (2012) Highly discriminative statistical features of email classification. *Knowl Inf Syst* 31(3):23–57
- Graham P (2003) Better Bayesian filtering. Talk at the 2003 spam conference. MIT
- Hutto CJ, Gilbert E (2014) VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: ICWSM. The AAAI Press
- Jiawei H, Kamber M (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann, Burlington, Massachusetts
- Koprinska I, Poon J, Clark J, Chan J (2007) Learning to classify e-mail. *Inf Sci* 177:2167–2187
- Lampert A, Dale R, Paris C (2009) Segmenting email message text into zones. In: Proceedings of the 2009 conference on empirical methods in natural language processing, Singapore, 6–7 Aug 2009, pp 919–928
- Li WJ, Hershkop S, Stolfo SJ (2004) Email archive analysis through graphical visualization. In: Proceedings of the 2004 ACM workshop on visualization and data mining for computer security (VizSEC/DMSEC'04). ACM, New York, pp 128–132
- Liu B (2006) Web data mining: exploring hyperlinks, contents, and usage data. Data-centric systems and applications. Springer, Berlin/Heidelberg
- Liu B (2010) Sentiment analysis and subjectivity. In: Indurkha N, Damerau F (eds) Handbook of natural language processing, 2nd edn. Chapman and Hall, Boca Raton
- Liu B (2012) Sentiment analysis and opinion mining. San Rafael, Morgan and Claypool
- McCallum A, Wang X, Corrada-Emmanuel A (2007) Topic and role discovery in social networks with experiments on enron and academic email. *J Artif Intell Res* 30(1):249–272
- Mimno D, McCallum A (2007) Expertise modeling for matching papers with reviewers. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'07). ACM, New York, pp 500–509
- Oard DW, Webber W (2013) Information retrieval for E-discovery. *Found Trends Inf Retr* 7(2–3):99–237
- Otte E, Rousseau R (2002) Social network analysis: a powerful strategy, also for the information sciences. *J Inf Sci* 28(6):441–453
- Pang B, Lee L (2008) Opinion mining and sentiment analysis. *Found Trends Inf Retr* 2(1):1–135
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), pp 79–86
- Pantel P, Lin D (1998) SpamCop – a spam classification & organization program. In: Proceedings of AAAI-98 workshop on learning for text categorization
- Radev DR, Jing H, Sty's M, Tam D (2004) Centroid-based summarization of multiple documents. *Inf Process Manag* 40:919–938
- Roitblat H, Kershaw A, Oot P (2010) Document categorization in legal electronic discovery: computer classification vs. manual review. *JASIST* 61:70–80
- Rowe R, Creamer G, Hershkop S, Stolfo SJ (2007) Automated social hierarchy detection through email network analysis. In: Proceedings of the 9th WebKDD and SNA-KDD 2007. ACM, New York, pp 109–117
- Sahami M, Dumais S, Heckerman D, Horvitz E (1998) A Bayesian approach to filtering junk E-Mail. In: Proceedings of AAAI-98 workshop on learning for text categorization
- Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. *Commun ACM* 18(11):613–620
- Schwartz MF, Wood DCM (1993) Discovering shared interests using graph analysis. *Commun ACM* 36(8):78–89
- Scott JP, Carrington PJ (2011) The SAGE handbook of social network analysis. Sage Publications Ltd., Newbury Park, California
- Tan PN, Steinback M, Kumar V (2006) Introduction to data mining. Pearson/Addison-Wesley, Boston, Massachusetts
- Tang G, Pei J, Luk WS (2014) Email mining: tasks, common techniques, and tools. *Knowl Inf Syst* 41:1
- Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in internet social networks. *J Mark Res* 47(4):643–658
- Turney P (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the association for computational linguistics, pp 417–424
- Turney PD, Littman ML (2003) Measuring praise and criticism. *ACM Trans Inf Syst* 21(4):315–346
- Tyler JR, Wilkinson DM, Huberman BA (2003) Email as spectroscopy: automated discovery of community structure within organizations. In: Huysman M, Wenger E, Wulf V (eds) Communities and technologies. Kluwer, B.V., Deventer, pp 81–96

- Vinjunur JK, Oard DW, Axelrod A (2016) An AID for avoiding inadvertent disclosure: supporting interactive review for privilege in E-discovery. In: ACM SIGIR conference on human information interaction and retrieval, Chapel Hill
- Witten IH, Frank E, Hall MA, Pal CJ (2016) Data mining: practical machine learning tools and techniques, 4th edn. Morgan Kaufmann, Burlington, Massachusetts
- Yeh JY, Harnly A (2006) Email thread reassembly using similarity matching. In: Proceedings of the third conference on email and anti-spam, CEAS 2006, Mountain View, 27–28 July 2006
- Yerazunis B (2003) Sparse binary polynomial hash message filtering and the CRM114 discriminator. In: Proceedings of 2003 spam conference. MIT

Fuzzy Similarity for Parallel Function Computation Model

Chin-Liang Chang
Nicesoft Corporation, Austin, TX, USA

Article Outline

Background
OpenCL (Open Computing Language)
Fuzzy Similarity for Natural Language Processing (NLP)
Translations in Natural Language Processing
Bibliography

Background

A neural net (of many layers in deep learning (Wikipedia)) is basically to represent a training set. However, some samples in the training set may not be classified correctly. A neural net is an expression-based approach, because the neural net may be considered as an expression.

We cite the following limitations of the neural net and deep learning approach:

- A brain has 86 billion neurons, and each with 1,000 connections. Thus, it is impossible to model a brain by a neural net.
- It is impossible to have a “deep learning” of a 20-layer neural net with 20 million samples in a training set.
- Risks with a damaged neural net and ill-trained neural net with wrong samples.
- A neural net is procedural (layer by layer). Therefore, parallel computation cannot be performed.
- Input and output of a neural net need to be numerical. That is, a neural net needs to work on numbers.
- A neural net is a black box. That is, we don’t know how weights in the neural net are

adjusted by input samples. Basically, a curve fitting method is used to develop the neural net.

On the other hand, we can consider the function computation model. In this case, an application is a function f , which is a set or a fuzzy set $\{x, f(x)\}$. For example, Chang (2015a, b), suppose we want to specify a fuzzy set “RED” f whose membership function is $f(r, b, g)$, where r , b , and g are integer values from 0 to 255. If we use a full size, then we need to specify $f(r, b, g)$ at $256 \times 256 \times 256$, or 16,777,216 points. This function table will be too big to handle. If we use the 8 grid size, we need to specify at $32 \times 32 \times 32$ or 32,768 points. If we use the 16 grid size, we need to specify at $16 \times 16 \times 16$ or 4096 points. We’ll use the 16 grid size. For each RGB value (r, g, b) of these 4096 points, we can display its color. By looking at the color, we can reliably give a grade of “RED.” The following are some colors, their RGB values, and their grades of “RED.”

	(255,15,0) 1.0		(239,15,0) 1.0		(223,15,0) 1.0
	(207,15,0) 1.0		(191,15,0) 1.0		(175,15,0) 0.9
	(159,15,0) 0.9		(143,15,0) 0.8		(127,15,0) 0.8
	(111,15,0) 0.7		(95,15,0) 0.5		(79,15,0) 0.4
	(63,15,0) 0.2		(47,15,0) 0.1		(31,15,0) 0.0
	(15,15,0) 0.0				

In Chang et al. (2019), we can see that an application is specified as a set, not necessarily a fuzzy set. In this case, matching (searching) can be done efficiently by indexing in a database system.

If we use “fuzzy similarity for parallel function computation model (Chang 2015a, b; Chang et al. 2019),” then we have the following advantages:

- Before we had a computer, we use a function table (e.g., sine function). If we want to find

$f(x)$, we try to find the most similar point m of x and take $f(m)$ as the value of $f(x)$.

- Parallel computation can be easily applied in this “best search” function computation model.
- A function f is a set $\{x, f(x)\}$ where if $x = y$ then $f(x) = f(y)$. It can be stored as a function table.
- To make the above approach accurate, the function table has to be large. That is why we need a “Massively Parallel Array Processor.” Fortunately, we now have multicore CPU and GPU (e.g., AMD’s Ryzen and EPYC, and Nvidia’s GPU with 5000 cores). These chips can be used as computing platforms to implement a massively parallel array processor.
- In our function computation model, we do not require that the domain and the range of function f to be numerical. They can be symbols. For example, the domain of f can be an image, a sentence, a product review, a Go position, etc., while the range of f can be a recognition class, an image description, an operation, a sequence of operations, a software program, etc.

In our approach to an application, we use a nearest (the most fuzzy similar) neighbor classifier to recognize patterns. First, we need to extract features from the patterns. For example, in facial image recognition, features may include distances between the eyes and distances between the eyes and the mouth. Note that the features can be computed in “parallel,” and each feature can be computed by a core in a GPUs computing platform, instead of one feature after another.

Given a training set $\{p_i \mid i = 1, \dots, m\}$, we choose every p_i , $i = 1, \dots, m$ to be a prototype in the application. Note that m can be large (e.g., 10,000 or 10 millions), and p_i can be a text, an image, a profile, a voice, a sequence of operations (Chang et al. 2019), etc.

Given an unknown pattern x , if x is represented by an n -dimensional vector of values of the same features, and if we use an Euclidean distance $d(x, p_i)$ between x and p_i , $i = 1, \dots, m$, to measure the nearness (fuzzy similarity) of the two vector, we can use this equality:

$$d(x, p_k) = \min \{ d(x, p_1), \dots, d(x, p_m) \},$$

and x will be classified as the class of p_k . In the above equality, we can replace $d(x, p_i)$ by $(p_i * p_i - 2 * x * p_i)$.

For example, let us consider the facial image recognition application in Schwartz (2018). We can choose passport images as our prototypes in a **training set**. Given an image x of a driver/passenger in a vehicle (taken by a camera) in a **test set**, the problem is to find the nearest (most fuzzy similar) prototype neighbor of x . The recognition rate is the percentage of those patterns in the test set, which are classified **correctly**.

The above formulation is quite natural. However, if we use the neural net formulation, it is “not” so straightforward.

In the neural net formulation, we need to decide the input and output of a neural net. The input of the neural net can be passport images or image x , and the output of the neural net is a class number. If the input is fed with the passport image of a person, or an image x of a driver/passenger in a vehicle (taken by a camera), then the output will be the class number assigned to him/her. Since this application handles many persons, the number of class numbers will be very large, and the neural net can be very hard to train. On the other hand, if the output is a profile (a text) of a person, instead of a class number of the person, it is difficult to represent the profile by a number for the neural net.

I agree with Microsoft that Congress should regulate facial recognition (Locklear 2018). Congress should require the software developers to publish methods they use and how do they test their products, their limitations, responsibilities, etc. I don’t know the reasons why so many big companies like Google, Facebook, Nvidia, etc. jump on “deep learning” and ignore the requirements that a neural net is a black box, and its input and output need to be numerical.

OpenCL (Open Computing Language)

OpenCL (Kaeli et al. 2015) is a framework for writing programs that execute across heterogeneous platforms consisting of central processing units (CPUs), graphics processing units (GPUs), digital signal processors (DSPs), field-programmable gate arrays (FPGAs), and other processors.

OpenCL specifies programming languages for programming these devices and application programming interfaces (APIs) to control the platform and execute programs on the compute devices. OpenCL provides a standard interface for **parallel computing** using task-based and data-based parallelism.

OpenCL is an open standard maintained by the nonprofit technology consortium Khronos Group. Conformant implementations are available from Altera, AMD, Apple, ARM, Creative, IBM, Imagination, Intel, Nvidia, Qualcomm, Samsung, Vivante, Xilinx, and ZiiLABS.

OpenCL Execution Model:

- **Compute Device:** A compute device is a collection of compute units. OpenCL compute devices typically correspond to a multicore GPU, a multicore CPU, or multicore DSP.
- **Compute Unit:** For multicore devices, a compute unit often corresponds to one of the cores. A work group executes on a single compute unit, and multiple work groups can execute concurrently on multiple compute units within a device. A compute unit will have local memory that is accessible only by the compute unit.
- **Command Queue:** A command queue is created for a specific device in a context. Command queues hold commands that will be executed on that specific device. A compute device may have many command queues associated with it, but a command queue will only associate with one compute device.
- **Kernel:** A kernel is a function declared in an OpenCL C program and executed on a compute device. Kernels are enqueued to compute devices through command queues.
- **Work item:** For `enqueueNDRangeKernel`, the number of work items is explicitly specified by the global size argument. A work item of a work group is executed on a compute unit.

A work item is distinguished from other executions within the collection by its global ID and local ID.

- **Work group:** A collection of related work items that execute on a single compute unit. The work items in the group execute the same kernel and share local memory.
- **Global ID:** A global ID is used to uniquely identify a work item and is derived from the number of global work items specified when executing a kernel.
- **Local ID:** A local ID specifies a unique work item ID within a given work group that is executing a kernel.

Chips with 7 nm or 5 nm technology (Bu 2017; Chacos 2015; Dent 2018; Hendersom 2014; Nield 2017) will be available in the near future. Any of these chips can hold more than 20 billion transistors. We have no doubt that we can get more than 5000 cores on a chip. Therefore, we can take the advantage of the parallelism in OpenCL to tackle **artificial intelligence applications with best mach parallel computation model**.

The following is a simple example of OpenCL with Microsoft Visual Studio C++. This example consists of two files, `proj1.cpp` and `proj1_Kernels.cl`, where `proj1.cpp` is a C++ program, and `proj1_Kernels.cl` is a kernel program which performs parallel computations on multicore computing platform.

`proj1.cpp` generates 1024 prototypes of 256-dimensional vectors and chooses the 21th prototype as the x-vector. It passes this information to the kernel program.

The kernel program, `proj1_Kernels.cl`, consists of two kernels. The first kernel computes the “fuzzy similarity” $(\pi_i * \pi_i - 2 * \pi_i * x_i)$, $i = 1, 1024$, where each vector of π_i and x is 256-dimensional vector. In the statements for “//Execute the kernel”

```
cl::NDRange global(1024);
cl::NDRange local(128);
```

it says that there are 1024 work items, and a work group consists of 128 work items. A computing unit (a core) is mapped to a work group.

The second kernel is to use a “reduce method” to find the minimum of

$(\pi * \pi - 2 * \pi * x), i = 1, 1024.$

proj1.cpp

// This program implements fuzzy similarity for function computation model using OpenCL

```
#define __CL_ENABLE_EXCEPTIONS
```

```
#include <CL/cl.hpp>
#include <iostream>
#include <fstream>
#include <string>
#include <vector>
```

```
using std::cout;
using std::cerr;
using std::endl;
using std::string;
```

```
void printVector(const std::string
arrayName,
                const float * arrayData,
                const unsigned int
length)
{
    int numElementsToPrint = (256 <
length) ? 256 : length;
    cout << endl << arrayName << ":"
<< endl;
    for(int i = 0;
i < numElementsToPrint; ++i)
        cout << arrayData[i] << " ";
    cout << endl;
}
```

```
void printVector1(const std::string
arrayName,
                const int * arrayData,
                const unsigned int
length)
{
    int numElementsToPrint = (256 <
length) ? 256 : length;
    cout << endl << arrayName << ":"
<< endl;
    for(int i = 0;
i < numElementsToPrint; ++i)
        cout << arrayData[i] << " ";
    cout << endl;
}
```

```
void getVector(const std::string
arrayName,
                const float * arrayData,
                const int dim,
                int row)
{
```

```
    cout << endl << arrayName << ":"
<< endl;
    int r = row*dim;
    for(int i = 0; i < dim; ++i)
        cout << arrayData[r + i] << "
";
    cout << endl;
}
```

```
int main() {
    const int dim = 256;
    const int elements = 1024;
    const int size = dim*elements;
    size_t datasize = sizeof(float)*
size;
    size_t vectorsize = sizeof
(float)*size;
    size_t samplesize = sizeof
(float)*elements;

    float *P = new float[size];
    float *X = new float[dim];
    float *C = new float[elements];
    float *scratch = new float
[elements];
    float *result = new float[elements];
    int *scratchIndex = new int
[elements];
    int *resultIndex = new int
[elements];

    for(int i = 0; i < elements; i++) {
        for(int j = 0; j < dim; j+
+) {
            P[i*dim + j] = (float) (i*dim +
j)/elements;
        }
    }

    for(int i = 0; i < dim; i++) {
        X[i] = (float) (20*dim + i)/
elements;
    }

    for(int i = 0; i < elements; i++) {
        scratchIndex[i] = i;
    }

    getVector("P", P, dim, 1);

    printVector("X", X, dim);

    try {
        // Query for platforms
        std::vector<cl::Platform>
platforms;
        cl::Platform::get(&platforms);
```

```

    // Get a list of devices on this
platform
    std::vector <cl::Device> devices;
    platforms[0].getDevices
(CL_DEVICE_TYPE_GPU, &devices);

    // Create a context for devices
    cl::Context context(devices);

    // Create a command-queue for the
first device
    //cl::CommandQueue queue =
cl::CommandQueue(context, devices[0]);
    cl::CommandQueue queue = cl::
CommandQueue(context,
                                devices[0],
CL_QUEUE_PROFILING_ENABLE);

    // Create the memory buffers
    cl::Buffer bufferP = cl::Buffer
(context, CL_MEM_READ_ONLY, datasize);
    cl::Buffer bufferX = cl::Buffer
(context, CL_MEM_READ_ONLY,
                                vectorsize);
    cl::Buffer bufferC = cl::Buffer
(context, CL_MEM_READ_WRITE,
samplesize);
    cl::Buffer bufferscratch = cl::
Buffer(context, CL_MEM_READ_WRITE,
samplesize);
    cl::Buffer bufferresult = cl::
Buffer(context, CL_MEM_READ_WRITE,
samplesize);
    cl::Buffer bufferscratchIndex =
cl::Buffer(context, CL_MEM_READ_WRITE,
elements*sizeof(int));
    cl::Buffer bufferresultIndex = cl::
Buffer(context, CL_MEM_READ_WRITE,
elements*sizeof(int));

    // Copy the input data to the input
buffers using the
// command-queue for the first device
    queue.enqueueWriteBuffer(bufferP,
CL_TRUE, 0, datasize, P);
    queue.enqueueWriteBuffer(bufferX,
CL_TRUE, 0, vectorsize, X);

    // Read the program source
    std::ifstream sourceFile
("proj1_Kernels.cl");
    std::string sourceCode(std::
istreambuf_iterator
< char > (sourceFile),
                                (std::
istreambuf_iterator < char > ());
    cl::Program::Sources source
(1, std::make_pair(sourceCode.c_str(),
sourceCode.length() + 1));

    // Create the program from the
source code
    cl::Program program = cl::Program
(context, source);

    // Build the program for the device
    cout << "aaaaaa" << endl;
    program.build(devices);
    cout << "bbbbbb" << endl;

    // Create the kernel
    cl::Kernel kernel_fuzzySimilarity
(program, "fuzzySimilarity");

    // Set the kernel arguments
    kernel_fuzzySimilarity.setArg
(0, bufferP);
    kernel_fuzzySimilarity.setArg
(1, bufferX);
    kernel_fuzzySimilarity.
setArg(2, dim);
    kernel_fuzzySimilarity.setArg
(3, bufferC);

    // Execute the kernel
    cl::NDRange global(1024);
    cl::NDRange local(128);
    queue.enqueueNDRangeKernel
(kernel_fuzzySimilarity, cl::NullRange,
global,
                                local);

    // Copy the putput data back to the
host
    queue.enqueueReadBuffer(bufferC,
CL_TRUE, 0, samplesize, C);
    printVector("C", C,
elements);

    queue.enqueueWriteBuffer
(bufferscratch, CL_TRUE, 0, samplesize,
C);
    queue.enqueueWriteBuffer
(bufferscratchIndex, CL_TRUE, 0,
samplesize,
                                scratchIndex);
    cl::Kernel kernel_reduce(program,
"reduce");
    kernel_reduce.setArg
(0, bufferscratch);
    kernel_reduce.setArg
(1, bufferresult);

```

```

        kernel_reduce.setArg
(2, bufferscratchIndex);
        kernel_reduce.setArg
(3, bufferresultIndex);

        cl::Event evt;
        queue.enqueueNDRangeKernel
(kernel_reduce, cl::NullRange, global,
local,
                                NULL, &evt);
        evt.wait();
        cl_ulong elapsed =
            evt.getProfiling
Info<CL_PROFILING_COMMAND_END>() -
            evt.get
ProfilingInfo<CL_PROFILING
_COMMAND_START>();
        cout << elapsed << endl;

        queue.enqueueReadBuffer
(bufferresult, CL_TRUE, 0, samplesize,
result);
        printVector("result", result,
elements);
        queue.enqueueReadBuffer
(bufferresultIndex, CL_TRUE, 0,
samplesize,
                                resultIndex);
        printVector1("resultIndex",
resultIndex, elements);
        cout << resultIndex[0] <<
endl;
        getVector("C", C, 1,
resultIndex[0]);

    }
    catch(cl::Error error)
    {
        std::cout << error.what() << "("
<< error.err() << ")" << std::endl;
    }
    return 0;
}

```

proj1_Kernels.cl

```

// Compute  $\Pi * \Pi - 2 * X * \Pi$  (float values
fuzzy similarity) for  $i = 0, \dots,$ 
elements
// Store results into C

__kernel
void fuzzySimilarity(__global float * P,
                    __global float * X,
                    int dim,
                    __global float * C)
{
    // Get the work-item's unique ID

```

```

    int idx = get_global_id(0);
    int row = idx*dim;
    int row1;

    // Compute  $\Pi * \Pi - 2 * X * \Pi$ , and store the
result in C
    float value1 = 0.0f;
    float value2 = 0.0f;
    for (int k = 0; k < dim; k++)
    {
        row1 = row + k;
        value1 += P[row1] * P[row1];
        value2 += X[k] * P[row1];
    }
    // Write the result to device memory
    C[idx] = value1 - 2*value2;
}

__kernel
void reduce(__global float* scratch,
            __global float* result,
            __global int* scratchIndex,
            __global int* resultIndex)
{
    int global_index = get_global_id(0);

    // Perform parallel reduction
    int local_index = get_local_id(0);
    // scratch[local_index] =
accumulator;
    barrier(CLK_LOCAL_MEM_FENCE);
    for(int offset = get_local_size(0) /
2;
        offset > 0;
        offset = offset / 2) {
        if (local_index < offset) {
            float other = scratch[local_index +
offset];
            float mine = scratch[local_index];
            //scratch[local_index] =
(mine < other) ? mine : other;
            if (mine < other) {
                scratch[local_index] = mine;
            } else {
                scratch[local_index] = other;
                scratchIndex[local_index] =
scratchIndex[local_index + offset];
            }
        }
        barrier(CLK_LOCAL_MEM_FENCE);
    }
    if (local_index == 0) {
        result[get_group_id(0)] = scratch
[0];
        resultIndex[get_group_id(0)] =
scratchIndex[0];
    }
}

```


Fuzzy Similarity for Natural Language Processing (NLP)

Introduction

Natural language processing covers many applications (e.g., text-to-text machine translation, text-to-voice machine translation, text-to-voice conversion, voice-to-text conversion, customer sentiment analysis, need-to-product/service matching, book cataloging, news reports cataloging, etc.) In this entry, we'll examine some present approaches and propose how fuzzy similarity can be used in natural language processing.

word2vec

Google made word2vec (Google) an open-source tool and posted it on the Internet in 2013.

word2vec uses deep learning and builds a large many-layer neural network. Coefficients in the neural network are trained on a huge database of text.

Using word2vec, one can get a 300-dimensional vector representation of any word. Word by word, word2vec tries to predict the other surrounding words in a sentence. It needs a huge database of example texts to train the vector of a word. Typically, this means hundreds of millions of words, which is the equivalent of 1,000 books, 500,000 comments, or 4,000,000 tweets. It may take hours or days to have the training done. Personally, I don't like it because it is not clear what is the meaning of the vector of a word. That is, what do axes in the 300-dimensional vector space represent? In addition, if I use "cosine similarity" to measure the similarity between any two words, I do not understand its meaning. For example, a tool based on word2vec, will give me similar (France,Spain) = 0.6, which is meaningless to me.

Mathematical Transformations

We repeat this section described in Chang (2015a, b). Basically, mathematical transformations are applied to a raw signal to obtain further information from that signal that is not readily available in the raw signal. Most of the signals in practice are time domain signals such as voices, EKG, radio signals, seismic signals, etc. Some popular mathematical transformations are Fourier transformation and wavelet

transformation. In this entry, we'll use an extended idea of mathematical transformations to extract **features**.

We choose a finite set of basis prototypes, $p_1, \dots, p_n$ and approximate a pattern p as a linear combination of these basis prototypes as

$$p = a_1 * p_1 + \dots + a_n * p_n.$$

However, in the above approximation, we don't require that $p_1, \dots, p_n$ are orthogonal. a_i is computed as a matching between p and p_i ,

$$a_i = p * p_i$$

where $i = 1, \dots, n$.

There are many ways to define $p * p_i$. It can be a normalized inner product representing cosine of the angle between p and p_i , or obtained by a hardware matching through a lens like an eye, or obtained by some fuzzy similarity matching.

How do we choose the basis prototypes? For pattern recognition, we may choose several patterns from each class, allowing many variants of patterns in each class. The variants may include enlarged, shrunk, translated patterns, etc. The basis prototypes may cover many different patterns we have collected.

Book Cataloging

As described in the following, the steps for book cataloging follow these steps:

1. Collect a set S of prototype books and their subject categories. There are already many existing books which have been assigned by librarians with known subject categories. Therefore, we do not need to repeat this step. We only need to select representative books. Set S could be large (e.g., one million books). That is why we want parallel computation described in Chap. 66 in Chang (2015a).
2. From S , choose a set B of n representative prototype books as basis prototype books. Each of B is considered as an axis in the n -dimensional Euclidean space.
3. For each p in S , represent it by an n -dimensional vector using B as the basis

prototypes. That is, computing a matching score p with every basis prototype in B . The matching scores will be the values in the n -dimensional vector. Let S' denote the set of n -dimensional vectors of all prototypes in S .

4. Given a book x , also compute an n -dimensional vector x' of x using B as the basis prototypes.
5. Find the nearest neighbor of x' in S' . The subject category of the nearest neighbor will be used as the subject category of x .

The above steps use the idea of the transformation method described previously to represent a book by an n -dimensional Euclidean vector. Then we use the Euclidean distance to find a nearest neighbor. I think that our n -dimensional vectors are more meaningful than a 300-dimensional vectors obtained by word2vec.

In Step 3, how do we match a prototype book p in S with every basis prototype book b in B ? We can select a set of keywords or phrases in p to match with another set of keywords or phrases in b . This is better illustrated by the following example.

I arbitrarily select books from Austin Public Library. Even though there is Library of Congress Classification (Library of Congress Classification), Austin Public Library uses Dewey Decimal Classification (DDC). The subject guide of DDC is given below:

000–009	General works Such as encyclopedias
100–199	Philosophy, psychology
200–299	Religion, mythology
300–399	Social sciences Government Sociology, economics Books on clubs Holiday and etiquette
400–499	Languages: Grammar Linguistics
500–599	Pure science Mathematics Physics, chemistry Botany, zoology, earth science Biological sciences
600–699	Applied science Inventions, gardening Cooking, sewing

	Aeronautics Engineering Medicine, agriculture
700–799	Arts and recreation Drawing, painting And music
800–899	Literature (except fiction)
900–999	History: Travel Geography, biography

To test our approach, we select the following prototype books:

p1: Book1

Title and year: Making a Difference (2012)

Authors: Captain Chesley “Sully” Sullenberger
With Douglas Century

Preface:

At a time of political polarization and economic turmoil, we yearn for superior leadership. Few have demonstrated this trait better than Captain “Sully” Sullenberger, a man who embodies the core values that are at the heart of America: responsibility, optimism, integrity, loyalty, and compassion. In this follow-up to his bestselling memoir, Sullenberger engages nearly a dozen distinguished Americans to explore the nature of leadership, what it means, what it takes, and how it can be fostered and developed in all of our lives.

Keywords/phrases: leadership, core value

DDC: 303.34

p2: Book2

Title and year: Green Illusions (2012)

Author: Ozzie Zehner

Preface:

We don't have an energy crisis. We have a consumption crisis. And this book, which takes aim at cherished assumptions regarding energy, offers refreshing straight talk about what's wrong with the way we think and talk about the problem. Though we generally believe we can solve environmental problem with more energy – more solar cells, wind turbines, and biofuels – alternative technologies come with their own side effects and limitations. How, for instance, do solar cells cause harm? Why can't engineers solve wind power's biggest obstacle? Why won't

contraception solve the problem of overpopulation at the heart of our concerns about energy, and what will?

Keywords/phrases: energy crisis, consumption crisis, renewable energy

DDC: 333.7940973

p3: Book3

Title and year: *How Babies Talk* (1999)

Authors: Roberta Michnick Golinkoff and Kathy Hirsh-Pasek

Preface:

Even before babies are born, nature and nurture are at work preparing them with the tools for mastering language. Almost magically, in the first 3 years of life, they learn to recognize words, decipher their meanings, put together sentences, and ask questions. Written by two experts in the field of language development, *How Babies Talk* explores how babies learn language and the concrete ways in which parents and caregivers can help nurture these linguistic skills at every stage of their children's development.

Keywords/phrases: learn language, language development

DDC: 401.93

p4: Book4

Title and year: *Kiss My Math* (2008)

Author: Danica McKellar

Preface:

With *Math Doesn't Suck*, actress and math genius Danica McKellar made wave nationwide, challenging the "math nerd" stereotype and giving students the tools to ace tests and homework. Now, in *Kiss My Math*, Danica empowers a new crop of girls, taking on the next level in the math curriculum: Pre-algebra.

Stepping up not only the math but also the sass and style, *Kiss My Math* will help math-phobic teenagers everywhere chill out and finally get negative numbers, variables, absolute values, exponents, graphing, and more – concepts essential for algebra and beyond.

Keywords/phrases: learn math, algebra

DDC: 512

p5: Book5

Title and year: *Plant Propagation* (1999)

Editor: Alan Toogood

Photographer: Peter Anderson

Preface:

The book gives the unrivaled practical guide to the successful propagation of all garden plants – from trees and shrubs to culinary herbs. It includes general and specialized techniques as well as unique tips from expert propagators.

Keyword/phrase: plant propagation

DDC: 635.043

p6: Book6

Title and year: *Hollywood Knits Style* (2007)

Author: Suss Cousins

Preface:

A passionate advocate of not only knitting but the good life in general, Cousins highlights her hospitality in *Hollywood Knits Style*. She shows knitters how to throw the best parties (a mother-daughter brunch, wedding shower, or girls' night out, all with a knitting theme), start a knitting circle, even fund-raising through knitting. To enhance the knitting experience, Cousins shares her favorite recipes for entertaining and her creative techniques for decorating a room with yarn.

Keyword/phrase: knitting

DDC: 746.432

p7: Book7

Title and year: *American Women Writers* (2011)

Editor: Elaine Showalter

Preface:

Inspired and informed by her groundbreaking history *A Jury of Her Peers*, Elaine Showalter's landmark anthology features the best work of writers ranging from Puritan poet Anne Bradstreet to contemporary starts like Annie Proulx and Jhumpa Lahiri.

For centuries, women have been marginalized and overlooked in American literary history, and Showalter's collection corrects this injustice, allowing us to see our famous women writers in their full literary context and to encounter scores of lesser known and forgotten writers who fully deserve to be rediscovered and enjoyed by new generations of readers.

Keyword/phrase: woman writer
DDC: 810.809287

p8: Book8

Title and year: A History of The World in 100 Objects (2010)

Author: Neil MacGregor

Preface:

When did people first start to wear jewelry or play music? When were cows domesticated and why do we feed their milk to our children? Where were the first cities and what made them succeed? Who developed math – or invented money?

The history of humanity is one of invention and innovation, as we have continually created new things to use, to admire, or to leave our mark on the world. In this groundbreaking book, Neil MacGregor, the director of the British Museum, charts a course through time and across civilizations to paint a compelling portrait of mankind's evolution, focusing on unexpected turning points.

Keywords/phrases: invention history, mankind evolution

DDC: 930.1

The following book is used as an input book x:

x: Bookx

Title and year: The Future: Six Drivers of Global Change (2013)

Author: AL Gore

Preface:

From the former vice president and #1 *New York Times* bestselling author comes An Inconvenient Truth for everything – a frank and clear-eyed assessment of the emerging forces that are reshaping our world and will continue to do so in the decades to come.

Our is time of revolutionary change that has no precedent in history. With the same passion he brought to the challenge of climate change, and with his decades of experience on the front lines of global policy, AL Gore surveys our planet's beclouded horizon and offers a sober, learned, and ultimately hopeful forecast in the visionary tradition of Alvin Toffler's *Future Shock* and John Naisbitt's *Megatrends*. In *the Future*, Gore

identifies the six critical drivers of global change: economic globalization, the Internet, global power shifting, flawed economic compass, genome, radical, and disruption of Earth's ecosystems.

Keywords/phrases: emerging force, revolutionary change

DDC: 303.4

We choose Book1, Book2, Book3, Book4, Book5, Book6, Book7, and Book8 as basis prototype books. That is, set B consists of these books. Now, we want to compute coordinate values p_1' , p_2' , p_3' , p_4' , p_5' , p_6' , p_7' , and p_8' and x' of p_1 , p_2 , p_3 , p_4 , p_5 , p_6 , p_7 , and p_8 and x , respectively, along these basis books by the method described below:

Given a basic book b and a prototype book p, let $c_1, \dots, c_r$ be keywords/phrases for b, and let $d_1, \dots, d_s$ be keywords/phrases for p.

We can list $c_1, \dots, c_r$ as

c_1
.
.
.
.
 c_r

Matching each c_i with $\{d_1, \dots, d_s\}$, where $i = 1, \dots, r$. Take the best match score

(i.e., max), using a similarity table. Sum these match scores for $i = 1, \dots, r$. The total value is used as a coordinate value of p along the b axis.

However, in order to avoid that the b axis dominate a result by a large number M of its keywords/phrases, we should normalize this sum by dividing it by M.

For our example, a similarity table contains the following similarity:

```
similarity(X,X) = 1.0
similarity(X,Y) = similarity(Y,X)
similarity("leadership",
"woman_writer") = 0.5
similarity("core_value",
"invention_history") = 0.4
similarity("core_value",
"mankind_evolution") = 0.6
```

```

similarity("learn_language",
"learn_math") = 0.5
similarity("language_development",
"mankind_evolution") = 0.6
similarity("energy_crisis",
revolutionary_change) = 0.8
similarity(consumption_crisis,
revolutionary_change) = 0.8

```

Using the basis prototype books and the above similarity table, we obtain the following coordinate vectors:

```

p1' = (2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.5, 1.0)
p2' = (0.0, 3.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)
p3' = (0.0, 0.0, 2.0, 0.5, 0.0, 0.0, 0.0, 0.6)
p4' = (0.0, 0.0, 0.5, 2.0, 0.0, 0.0, 0.0, 0.0)
p5' = (0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0)
p6' = (0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0)
p7' = (0.5, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0)
p8' = (1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 2.0)
x' = (0.0, 1.6, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)

```

Since $p2'$ is the nearest neighbor of x' , we take the DDC of $p2$, 333.7940973, as the DDC of x . Or at least, 333 can be a good subject category for Bookx.

Of course, this is just a simple example to illustrate our method. In the real world, we may have chosen a large number of prototype books (e.g., one million prototype books) and 100 basis prototype books. In addition, the number of keywords/phrases associated with each book may be more than 2 or 3. In this case, the computing effort will be tremendous. Therefore, we need parallel processors.

There are also comments of specifying a similarity table for a big number of prototype books. I advocate specifying it by experienced persons.

We could use "crowd source" to have many persons to divide up the job. For example, Microsoft collects data from many players for its Kinect system, and Google has users uploaded videos to its YouTube website

It is hard to find a method for finding a good similarity table. As I have said, word2vec will not do the job.

Even an ontology-based method will not do the job either. For example, if "similarity" is defined by the ontology-based method, then similarity ("tiny," "small") should be larger than similarity ("tiny," "large") assigned by ontology *automatically*.

Translations in Natural Language Processing

One of the goals in Japanese Fifth Generation Computing Project in the 1980s was to talk to a phone in one language and to hear it in another language. This goal was a language translation application. The project was based on prolog. However, it was impossible to have "parallel" computation in prolog. Therefore, the goal failed.

We model the language translation application as a function, which is a set $\{s, f(s)\}$, where s is a sentence and $f(s)$ is its translation.

The Liberty Times news has published more than 500 Hanji-English translation articles. I have collected these 500 articles (Chang). Some titles of these articles are shown as follows:

- A village in New Zealand is trying to ban all cats (2018-09-04)
- Africa's largest mammal is terrified of this tiny insect (2018-09-03)
- Starbucks Korea to test paper straws (2018-09-02)
- New Zealand to ban plastic shopping bags from next year (2018-08-29)
- Scientists claim viagra could new miracle cure for blindness (2018-08-28)
- Tourists go off the beaten path on North Korea's sacred volcano (2018-08-26)
- Taiwan placed to benefit from transition to 5G mobile (2018-08-21)
- DeepMind created an IQ test for AI, and it didn't do too well DeepMind (2018-08-20)

- Paris residents peeved at very public eco-friendly urinals (2018-08-18)
- Immigration crackdown in Guangzhou's "Little Africa" (2018-08-16)
- Government to install solar power equipment in graveyards (2018-08-14)
- Christmas comes early for London department store (2018-08-11)
- Yale study: people who read live longer (2018-08-07)
- Colombian drug traffickers put a \$70,000 bounty on a police dog (2018-08-06)
- Underground lake found on Mars, raising possibility of life (2018-08-05)
- In wealthy Germany, growing up poor is a "dead-end street" (2018-08-02)
- Formosan black bear cub separated from mother to be treated and released back into wild (2018-08-01)
- Cyclist carries injured stray dog on his back, finds pup a forever home (2018-08-01)
- Teens who are constantly on their phones may be at risk of ADHD (2018-07-31)
- Taiwan's "green-collar" economy trumpeted as new gold rush (2018-07-31)
- La Paz's colorful cable car system (2018-07-30)
- Over 5000 score "zero" for English composition in joint exam (2018-07-26)
- Muslims in China's "Little Mecca" fear eradication of Islam (2018-07-26)
- India trains delayed due to "drunk" station master (2018-07-25)
- Moon hopes to declare end to Korean War this year; experts say not impossible (2018-07-22)
- World's oldest bread found at prehistoric site in Jordan (2018-07-22)
- Minnesota online sensation raccoon captured atop skyscraper (2018-07-21)
- Jingle bells: Japan's unusual station music (2018-07-19)
- Not just land heat waves: Oceans are in hot water, too (2018-05-19)
- Eating bananas could prevent heart attacks and strokes (2018-05-15)

Each article has about ten sentences. Therefore, the collection has about 5000 sentences. Using my model, the collection is a set $\{s, f(s)\}$, where s is an English sentence, and $f(s)$ is its Hanji (Chinese) translation. Given a new English sentence x , we want to find an English sentence m in the collection such that (x, m) has the most "fuzzy similar" match and use the translation of m as the example for the translation of x .

Note that the collection can be enlarged as the time goes by.

Bibliography

- Bu H (2017) 5 nanometer transistors inching their way into chips. IBM Research, 5 June 2017. <https://www.ibm.com/blogs/think/2017/06/5-nanometer-transistors/>
- Chacos B (2015) IBM's crazy-thin 7nm chip will hold 20 billion transistors, 9 July 2015. <https://www.pcworld.com/article/2946124/ibm-reveals-worlds-first-working-7nm-processor.html>
- Chang C-L (1974) Finding prototypes for nearest neighbor classifiers. *IEEE Trans Comput* C-23(11): 1179–1184
- Chang C-L (2015a) Applications of Fuzzy Similarity and My Internet Products, Version 1.0. Nicesoft Corporation, Austin
- Chang C-L (2015b) Fuzzy similarity for function computation model. In: Proceedings of the 34th annual conference of NAFIPS (North American Fuzzy Information Processing Society) and anniversary of Lotfi A. Zadeh's first paper on "Fuzzy Sets", Redmond, Washington, 17–19 Aug 2015
- Chang C-L Examples of English-to-Hanji and Hanji-to-English translations. <http://www.nicesoftsearch.com/findBlogs2.php?business=english-to-hanji>
- Chang C-L, Chang M, Leong P (2019) Solving sliding tile puzzle games by exact match. 2019 applications of fuzzy similarity workshop, Austin
- Dent S (2018) AMD is gearing up for 7-nanometer CPUs and graphic cards, 30 April 2018. <https://www.engadget.com/2018/04/30/amd-7-nanometer-cpus-and-gpus/>
- Google. word2vector Tools. <https://code.google.com/p/word2vec/>
- Hendersom R (2014) Intel claims that by 2026 processors will have as many transistors as there are (86 billion) neurons in a brain, 6 Jan 2014. <https://www.pocket-lint.com/laptops/news/intel/126289-intel-claims-that-by-2026-processors-will-have-as-many-transistors-as-there-are-neurons-in-a-brain>
- Kaeli D, Mistry P, Schaa D, Zhang DP (2015) Heterogeneous computing with OpenCL 2.0. Morgan Kaufmann/Elsevier, Waltham

- Library of Congress Classification. https://en.wikipedia.org/wiki/Library_of_Congress_Classification
- Locklear M (2018) Microsoft calls on congress to regulate facial recognition, 13 July 2018. <https://www.engadget.com/2018/07/13/microsoft-congress-regulate-facial-recognition/>
- Nield D (2017) IBM's new computer chips can fit 30 billion transistors on your fingertips, 6 June 2017. <https://www.sciencealert.com/new-computer-chips-can-fit-30-million-transistors-on-your-fingertip>
- Schwartz J (2018) Facial recognition will get border test, 6 Feb 2018. <http://www.nicesoftware.com/face.html>
- Wikipedia. Deep learning. https://en.wikipedia.org/wiki/Deep_learning

Mereology

Hsing-chien Tsai
Department of Philosophy, National Chung-
Cheng University, Chia-Yi, Taiwan

Article Outline

Glossary
Introduction
First-Order Logic
First-Order Mereological Axioms and Theories
Alternative Formulations
Models of Mereological Theories
Decidability Concerning First-Order
Mereological Theories
Fuzzification of Mereology
Future Directions
Bibliography

Glossary

Boolean algebra A Boolean algebra is a structure of the form $(A, \oplus, \otimes, \sim, 0, 1)$, where A is a nonempty set, $\oplus$ and $\otimes$ are two binary functions on A , $\sim$ is a unary function on A , and 0 and 1 are two distinct members in A such that both $\oplus$ and $\otimes$ are commutative and associative, $\oplus$ is distributive over $\otimes$, $\otimes$ is distributive over $\oplus$, $(x \oplus y) \otimes y = y$ and $(x \otimes y) \oplus y = y$ (absorption laws), and $x \oplus \sim x = 1$ and $x \otimes \sim x = 0$. By defining $x \leq y$ as $x \oplus y = y$, we can get a partial ordering from a Boolean algebra. A Boolean algebra $(A, \oplus, \otimes, \sim, 0, 1)$ is complete if and only if for any nonempty $S \subseteq A$, there is a member x in A such that for any $y \in S$, $y \leq x$ and for any z in A , if for any $y \in S$, $y \leq z$, then $x \leq z$ (x is called the least upper bound of S). Given a Boolean algebra, if we remove 0 from it, the resultant structure is called a Boolean algebra with 0 removed.

Decidability A set S of sentences in a formal language is decidable if and only if there is an effective procedure such that given any sentence α in that language, it can be answered by using such a procedure whether or not $\alpha \in S$.

Parthood relation The parthood relation is the binary relation “being a part of.” It is usually suggested that any formal interpretation of such a relation should be a partial ordering.

Partial ordering A partial ordering is a structure of the form $(A, \leq)$, where A is a nonempty set and $\leq$ is a binary relation on A such that $\leq$ is reflexive, transitive, and anti-symmetric. For any partial ordering, define $x < y$ if and only if $x \leq y$ and $x \neq y$. Then $(A, <)$ is called a strict partial ordering, in which $<$ is irreflexive and transitive. Given any strict partial ordering $(A, <)$, we can also get a partial ordering $(A, \leq)$ by defining $x \leq y$ if and only if $x < y$ or $x = y$.

Russell’s paradox Consider the set $R = \{x: x \notin x\}$. Then $R \in R$ if and only if $R \notin R$, which leads to a contradiction. It is generally thought that the problem originates from a naïve set-theoretical principle that for any predicate $\alpha(x)$, there is a set consisting of exactly all those members each of which satisfies $\alpha(x)$.

Introduction

Mereology, whose etymology is from the Greek μέρος, meaning “part,” is the theory of “the relations of part to whole and the relations of part to part within a whole” (Varzi 2016). Although the study of mereology dates back to Presocratics and continues through to the nineteenth century with many enlightening philosophical writings, mereology was never clearly formulated until the early twentieth century, and it was the Polish logician Stanisław Leśniewski who gave the first clear formulation of mereology (Leśniewski 1916).

Leśniewski’s theory contains the following two postulates (this is a rephrased version given in Tarski 1956a):

Postulate I. The relation “being a part of,” or the “parthood” relation, is transitive, that is, if x is a part of y and y is a part of z , then x is a part of z .

Postulate II. For any nonempty class S of members, there is exactly one member x which is a sum of all members in S , that is, each member in S is a part of x and every part of x overlaps some member in S , where overlapping means sharing at least one part.

We can easily find some mathematical structures which satisfy the foregoing two postulates. For instance, let A be the set of all nonempty subsets of the set of natural numbers, then the structure $(A, \subseteq)$, that is, the structure whose domain is A and whose binary relation is the set inclusion, satisfies the aforementioned two postulates (it is trivial that $\subseteq$ is transitive and for any nonempty $S \subseteq A$, $\cup S$ is the sum of all members in S). It is not difficult to see that $(A, \subseteq)$ is in effect a complete Boolean algebra with 0 removed. More generally, a structure satisfies postulates I and II if and only if it is a complete Boolean algebra with 0 removed (Grzegorczyk 1955; Tarski, 1956b; Clay 1974).

Postulate II is second-order since it involves quantifying over subsets of the domain. Moreover, Leśniewski’s theory is not first-order axiomatizable. Let B be the set of all nonempty finite subsets and all cofinite subsets of the set of natural numbers. Then it can be shown that $(A, \subseteq)$ and $(B, \subseteq)$ are elementarily equivalent (i.e., they satisfy the same first-order sentences) (Pietruszczak 2015; Tsai 2015), but it is obvious that $(B, \subseteq)$ does not satisfy Postulate II.

It is said that Leśniewski was shocked by Russell’s paradox which haunted the naïve set theory at that time. This urged him to look for another theory which can serve as an alternative foundation of mathematics, and the outcome was his mereology. Leśniewski’s theory won’t be plagued by Russell’s paradox, for it is easy to show that his theory implies that everything is a part of itself, and therefore, there is nothing which can be formed by those things each of which is not a part of itself (Postulate II applies to nonempty

classes only). A mathematical theory may serve as a foundation of mathematics if it suffices to define all kinds of mathematical structures (for instance, it is known that using the first-order axiomatized set theory **ZFC**, we can define all kinds of mathematical structures (see Kunen 1980)). However, Leśniewski’s theory does not seem to be strong enough to serve as a foundation of mathematics, for it is just something equivalent to the theory of complete Boolean algebras (see above). As we will see, the first-order approximation of Leśniewski’s theory is called general extensional mereology, **GEM**, which is a decidable theory. However, it is known that if a first-order theory suffices to define a certain kind of weak arithmetic theory, then such a theory must be undecidable (see Chap. 3 in Enderton 2001). This shows that **GEM** cannot define the arithmetic structure.

First-Order Logic

We will be mainly concerned with first-order logic in the following and here a brief survey (a rephrased version based on Enderton 2001) of what it is should be in order. A first-order language can use two kinds of symbols.

Logical symbols Parentheses: $(,)$; connectives: $\neg, \rightarrow$; quantifier: $\forall$; variables (enumerably infinitely many): $v_1, v_2, \dots$; equality symbol (optional): $=$.

The intended meanings of those symbols are as follows: “ $\neg$ ” means “not,” which is a monadic connective; “ $\rightarrow$ ” means “if...then,” which is a binary connective; and “ $\forall$ ” means “for all.” The meanings of parentheses and the equality symbol go without saying.

Parameters Predicate symbols (optional); function symbols (optional); and constant symbols (optional). Note that the arity of a predicate symbol or a function symbol must be greater than 0 .

Different first-order languages will differ in the parameters they have or in whether the equality symbol is used. When we are defining a first-order language, its parameters will be defined explicitly at the beginning and will then be fixed henceforth, and, of course, whether or not the equality symbol is used in the language will also be specified. Let

$\text{Log}(L)$ stand for the set of logical symbols used in L and $\text{Sig}(L)$ stand for the set of parameters used in L . An “expression” of a first-order language L is a finite sequence formed by using symbols in $\text{Log}(L) \cup \text{Sig}(L)$. Then “terms” as well as “formulas” will be formed in an inductive way as follows.

Terms Call a set S of expressions of L “inductive in the first sense” if (i) all variables and constant symbols (each of them is viewed as a sequence of length one) in $\text{Log}(L) \cup \text{Sig}(L)$ are in S and (ii) for any n -ary function symbol F in $\text{Sig}(L)$, any expressions $t_1, \dots, t_n$ in S , $F(t_1, \dots, t_n)$ is also in S . Then the set of terms is defined as the smallest set which is “inductive in the first sense” or the intersection of all sets each of which is “inductive in the first sense.” Let $\text{Term}(L)$ stand for the set of terms of L .

Formulas Similarly, call a set S of expressions of L “inductive in the second sense” if (i) for any n -ary predicate symbols P in $\text{Log}(L) \cup \text{Sig}(L)$, any n terms $t_1, \dots, t_n$ in $\text{Term}(L)$, $P(t_1, \dots, t_n)$ is in S and (ii) for any expressions α and β in S , $(\neg\alpha)$, $(\alpha \rightarrow \beta)$, and $\forall x\alpha$ are also in S . Then the set of formulas is defined as the smallest set which is “inductive in the second sense.” Let $\text{Formula}(L)$ stand for the set of formulas of L . Note that we will put $\neg\alpha$ and $\alpha \rightarrow \beta$ instead of $(\neg\alpha)$ and $(\alpha \rightarrow \beta)$ if no confusion should occur.

Atomic formula A formula α is atomic if neither connectives nor quantifiers occur in α . So an atomic formula must be some $P(t_1, \dots, t_n)$, where P is an n -ary predicate symbol and $t_1, \dots, t_n$ are terms.

Free and bound variables A variable in a formula is “free” if it is not within the scope of any quantifier in that formula. Otherwise, it is “bound.”

Sentence A sentence is a formula which does not have any free variable. For example, $\forall x(R(x) \rightarrow Q(x))$ is a sentence, but $\forall xR(x) \rightarrow Q(x)$ is not, for the second x is a free variable. Note that we will use x, y , and z as variables without specifying which v_i 's they are (recall that variables are those $v_1, v_2, v_3, \dots$).

Defining other symbols Suppose α and β are two formulas. Let “ $\alpha \wedge \beta$ ” be an abbreviation of “ $\neg(\alpha \rightarrow \neg\beta)$,” where the intended meaning of “ $\wedge$ ” is “and.” Let “ $\alpha \vee \beta$ ” be an abbreviation of

“ $\neg\alpha \rightarrow \beta$,” where the intended meaning of “ $\vee$ ” is “or.” Let “ $\alpha \leftrightarrow \beta$ ” be an abbreviation of “ $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$,” that is, “ $\neg((\alpha \rightarrow \beta) \rightarrow \neg(\beta \rightarrow \alpha))$,” where the intended meaning of “ $\leftrightarrow$ ” is “if and only if.” Finally, let “ $\exists x\alpha$ ” be an abbreviation of “ $\neg\forall x\neg\alpha$,” where the intended meaning of “ $\exists$ ” is “for some.”

The foregoing gives a quick survey of the syntactics of first-order logic. The semantics of first-order logic consists of two parts: the first is the definition of a “structure” or an “interpretation,” and the second is the definition of “satisfaction.”

Structure or interpretation Consider a first-order language L . A structure of L , or an L -structure, or an interpretation of L is of the form (U, I) , where U is a nonempty set and I is a mapping whose domain is $\text{Sig}(L)$ such that for any constant symbol $d \in \text{Sig}(L)$, $I(d) \in U$; for any n -ary predicate symbol $P \in \text{Sig}(L)$, $I(P) \subseteq U^n$, where $U^n = \{S: S \text{ is an } n \text{ tuple } (x_1, \dots, x_n), \text{ where each } x_i \in U, \text{ for } 1 \leq i \leq n\}$; and for any n -ary function symbol $F \in \text{Sig}(L)$, $I(F)$ is a function from U^n to U .

Assignment function Given a structure $M = (U, I)$, an assignment function g on M is a mapping from the set of variables to U . A given assignment function g can be extended to a function g' from $\text{Term}(L)$ to U in the following recursive way. If t is a variable, $g'(t) = g(t)$; if t is a constant, $g'(t) = I(t)$; if $t = F(t_1, \dots, t_n)$, where F is an n -ary function and $t_1, \dots, t_n$ are terms, $g'(t) = I(F)(g'(t_1), \dots, g'(t_n))$. It can be shown that g' is unique.

Satisfaction Suppose L is a first-order language. Let $M = (U, I)$ be an L -structure and g be an assignment function on M . For any $\alpha \in \text{Formula}(L)$, “ M satisfies α under g ” is symbolized as “ $M \models \alpha[g]$ ” and is defined inductively as follows.

If α is atomic, that is, $\alpha = P(t_1, \dots, t_n)$, where P is an n -ary predicate symbol and $t_1, \dots, t_n$ are terms, then $M \models \alpha[g]$ if and only if $(g'(t_1), \dots, g'(t_n)) \in I(P)$.

If $\alpha = \neg\beta$, then $M \models \alpha[g]$ if and only if not $M \models \beta[g]$.

If $\alpha = \beta \rightarrow \gamma$, then $M \models \alpha[g]$ if and only if either not $M \models \beta[g]$ or $M \models \gamma[g]$.

If $\alpha = \forall x\beta$, then $M \models \alpha[g]$ if and only if for any $d \in U$, $M \models \beta[g_{d,x}]$, where $g_{d,x}$ is an assignment function defined as follows: $g_{d,x}(x) = d$ and for any variable v other than x , $g_{d,x}(v) = g(v)$.

Note that if α is an equation, say, $u = t$, where u and t are terms, then $M \models \alpha[g]$ if and only if $g'(u) = g'(t)$. This cannot be changed, that is, the interpretation of the equality symbol is always the identity relation. Moreover, it can be shown that if α is a sentence, $M \models \alpha[g]$, for some assignment function g if and only if $M \models \alpha[g]$, for all assignment function g . So the reference to g can be dropped if the satisfaction is about a sentence.

Since only finitely many variables can occur in a formula, sometimes we will write $M \models \alpha(x_1, \dots, x_n)$ [$a_1, \dots, a_n$], which means M satisfies $\alpha(x_1, \dots, x_n)$ under the finite sequence $[a_1, \dots, a_n]$, which can be thought of as a finite assignment function f from $x_1, \dots, x_n$ to $a_1, \dots, a_n$ such that $f(x_i) = a_i$, for $1 \leq i \leq n$ (normally, it is assumed that $x_1, \dots, x_n$ are free in α when we are using the aforementioned notation).

Logical implication Given a first-order language L , a set $S \subseteq \text{Formula}(L)$ logically implies some $\alpha \in \text{Formula}(L)$ if and only if for any L -structure M , any assignment function g on M , if for each $\beta \in S$, $M \models \beta[g]$, then $M \models \alpha[g]$. Again, the reference to g can be dropped if all formulas involved are sentences.

Logical theory A set of sentences in a formal language is called a logical theory if such a set is closed under logical implication. For any set S of sentences, let $\text{Th}(S)$ stand for the closure of S under logical implication, and then by definition, $\text{Th}(S)$ is a logical theory. A logical theory is first-order if its formal language is first-order.

Complete theory A logical theory T is complete if and only if for any sentence α in the language of T , either $\alpha \in T$ or $\neg\alpha \in T$.

Elementary equivalence Given a first-order language L , two L -structures M_1 and M_2 are elementarily equivalent if and only if they satisfy the same sentences of L , that is, for any sentence α in L , $M_1 \models \alpha$ if and only if $M_2 \models \alpha$.

Recursive axiomatizability A logical theory T in a formal language L is recursively axiomatizable if and only if there is a decidable set S of sentences in L such that $T = \text{Th}(S)$. A logical theory T in a formal language L is

finitely axiomatizable if and only if there is a finite set S of sentences in L such that $T = \text{Th}(S)$. It should be clear that a finite set must be decidable.

Note that variables in a first-order language are called “individual variables,” each of which will denote a member in the domain, but a second-order language will also have predicate variables, for example, monadic predicate variables, each of which will denote a subset of the domain. Recall that Leśniewski’s Postulate II talks about “any nonempty class S of members,” and therefore its formalization will involve at least a monadic predicate variable, which is second-order.

First-Order Mereological Axioms and Theories

Leśniewski’s theory is called “classical mereology.” But there are still other mereologies or rather mereological theories which have been set up by more recent authors. An approach that can be found in the recent literature is to formulate mereological axioms first and then put some axioms together (if they are consistent) to form a mereological theory (the formulations and nomenclatures to be seen in the following mainly follow Simons 1987 and Casati and Varzi 1999).

The first-order language of mereology has only one parameter, a binary predicate P , whose intended meaning is the relation “being a part of” or the “parthood” relation. The equality symbol will also be used by such a language. In order to shorten complicated expressions, three additional predicates are defined as follows.

Proper Part: $PPxy =_{df} Pxy \wedge \neg Pyx$

Overlap: $Oxy =_{df} \exists z(Pzx \wedge Pzy)$

Underlap: $Uxy =_{df} \exists z(Pxz \wedge Pzy)$

First we have the three most basic mereological axioms.

$$(P1) \forall x Pxx$$

$$(P2) \forall x \forall y ((Pxy \wedge Pyx) \rightarrow x = y)$$

$$(P3) \forall x \forall y \forall z ((Pxy \wedge Pyz) \rightarrow Pxz)$$

(P1) is reflexivity, (P2) is anti-symmetry, and (P3) is transitivity. The theory axiomatized by

these three basic axioms is called ground mereology **GM** (this can be conveniently expressed by putting **GM** = (P1) + (P2) + (P3), and henceforth we will use the same kind of notation when defining a theory). We will require that any mereological theory must contain **GM** or, in other words, must have (P1), (P2), and (P3).

The following axioms can also be found in the literature: extensionality principle (EP), proper parts principle (PPP), weak supplementation principle (WSP), strong supplementation principle (SSP), finite sum (FS), and finite product (FP).

$$(EP) \forall x \forall y (\exists z PPzx \rightarrow (\forall z (PPzx \leftrightarrow PPzy) \rightarrow x = y))$$

$$(PPP) \forall x \forall y (\exists z PPzx \rightarrow (\forall z (PPzx \rightarrow PPzy) \rightarrow Pxy))$$

$$(WSP) \forall x \forall y (PPxy \rightarrow \exists z (PPzy \wedge \neg Ozx))$$

$$(SSP) \forall x \forall y (\neg Pyx \rightarrow \exists z (Pzy \wedge \neg Ozx))$$

$$(FS) \forall x \forall y (Uxy \rightarrow \exists z \forall w (Owz \leftrightarrow (Owx \vee Ow y)))$$

$$(FP) \forall x \forall y (Oxy \rightarrow \exists z \forall w (Pwz \leftrightarrow (Pwx \wedge Pwy)))$$

There is also a first-order approximation of Leśniewski's Postulate II, that is, unrestricted fusion (UF).

$$(UF) \exists x \alpha \rightarrow \exists z \forall y (Oyz \leftrightarrow \exists x (\alpha \wedge Oyx)),$$

for any formula α
where z and y do not occur free.

Note that (UF) is an axiom schema which defines simultaneously infinitely many axioms (α is a meta-variable which ranges over formulas). Moreover, every axiom should be a sentence. But α may have free variables other than x . For example, suppose $y_1, y_2, \dots, y_n$ occur free in α and each of them is distinct from x . In that case, the axiom generated by using (UF) will be $\forall y_1 \forall y_2 \dots \forall y_n (\exists x \alpha \rightarrow \exists z \forall y (Oyz \leftrightarrow \exists x (\alpha \wedge Oyx)))$.

The following mereological theories can be found in the literature.

Minimal Mereology **MM** = **GM** + (WSP)

Extensional Mereology **EM** = **GM** + (SSP)

Closure Mereology **CM** = **GM** + (FS) + (FP)

Minimal Extensional Mereology
MEM = **MM** + (FP)

Minimal Closure Mereology
CMM = **MM** + (FS) + (FP)

Extensional Closure Mereology
CEM = **EM** + (FS) + (FP)

General Extensional Mereology
GEM = **EM** + (UF)

It is obvious that all theories listed above except **GEM** are finitely axiomatizable. However, **GEM** is still recursively axiomatizable since there is an effective procedure via which we can check whether a formula is an instance of (UF) or not. More generally, if a first-order language can be "recursively numbered," there will be an effective procedure to check its formulas (see Chap. 3 in Enderton 2001).

Some notable orderings of mereological theories are as follows (Tsai, 2009).

$$\mathbf{GM} \subset \mathbf{MM} \subset \mathbf{EM} \subset \mathbf{MEM} \subset \mathbf{CMM} = \mathbf{CEM} \subset \mathbf{GEM}$$

$$\mathbf{GM} + (EP) \subset \mathbf{GM} + (PPP) \subset \mathbf{MM} + (PPP) \subset \mathbf{EM}$$

GM + (EP) and **MM** are independent of each other, that is, **GM** + (EP) $\neq$ **MM** and neither properly includes the other. **GM** + (PPP) does not include **MM**, and therefore, they are independent of each other. Furthermore, it is also not difficult to see that **CM** is independent of any of **EM**, **MM**, **GM** + (EP), and **GM** + (PPP).

There are also other axioms available such as complementation (C), greatest member (G), atomicity (A), and atomlessness (AL).

$$(C) \forall x (\neg \forall z Pzx \rightarrow \exists z \forall w (Pwz \leftrightarrow \neg Ow x))$$

$$(G) \exists x \forall y Pyx$$

$$\forall x \exists y (Pyx \wedge \neg \exists z PPzy)$$

$$(AL) \forall x \exists y PPyx$$

Both (C) and (G) are theorems of **GEM**, but neither (A) nor (AL) is. As a matter of fact, all axioms listed so far except (A) and (AL) are provable from **GEM**. Both **GEM** + (A) and **GEM** + (AL) are consistent, for the former is satisfied by any partial ordering with only one

member, for instance, $(\{\emptyset\}, \subseteq)$, and the latter is satisfied by any atomless Boolean algebra with 0 removed.

As mentioned earlier, a structure satisfies Leśniewski's theory if and only if it is a complete Boolean algebra with 0 removed. Similarly, it can be shown that a structure satisfies **CEM** + (C) + (G) if and only if it is a Boolean algebra with 0 removed (see section “[Models of Mereological Theories](#)” below).

Leśniewski's theory is strictly stronger than **GEM** in the sense that the structure $(B, \subseteq)$ defined earlier satisfies **GEM** but does not satisfy Leśniewski's theory. Intuitively, **GEM** should be strictly stronger than **CEM** + (C) + (G), but the structure which satisfies the latter but not the former is much more complicated. Such a structure must have infinitely many atoms (an atom is a member which does not have a proper part) as well as atomless members (a member is atomless, or a piece of “gunk,” if it contains no atom, and a structure must have infinitely many atomless members if it has at least one; see section “[Models of Mereological Theories](#)” below).

Because of (UF), **GEM** contains infinitely many axioms. However, it turns out that **GEM** is finitely axiomatizable. Let's define two monadic predicates as follows.

$$\text{Atom}(x) = \forall y \neg PPyx$$

$$\text{Gunk}(x) = \forall y (Pyx \rightarrow \exists z PPyz)$$

Then **GEM** is equivalent to **CEM** + (C) + (G) + $\exists x \text{Atom}(x) \rightarrow \exists z \forall y (Oyz \leftrightarrow \exists x (\text{Atom}(x) \wedge Oyx))$ or to **CEM** + (C) + (G) + $\exists x \text{Gunk}(x) \rightarrow \exists z \forall y (Oyz \leftrightarrow \exists x (\text{Gunk}(x) \wedge Oyx))$ (The fact that **GEM** is finitely axiomatizable has first been shown by Pietruszczak (2000), then by Niebergall (2014), and finally by the present writer (2018), but those writers worked out the same result independently through different approaches, and they gave different, though equivalent, finite axiomatizations.).

Alternative Formulations

In the foregoing, the binary predicate P is the sole primitive of the formal language of mereology, and other predicates can be defined by using P. It is possible to start with another predicate to develop mereological axioms and theories. Three alternatives can be seen in the literature. The first is to take a binary predicate $\text{NO}(x, y)$, whose intended meaning is “not overlapping,” as the primitive (see (Leonard and Goodman 1940), where their primitive is “discrete,” but according to their definition, two individuals are discrete if and only if they do not share any part). Then predicate P can be defined as follows.

$$Pxy =_{\text{df}} \forall z (\text{NO}(z, y) \rightarrow \text{NO}(z, x))$$

The definition seems to be intuitive. Oxy is still defined as $\exists z (Pzx \wedge Pzy)$. It is not surprising that the following axiom is desirable (Leonard and Goodman 1940).

$$\forall x \forall y (\text{NO}(x, y) \leftrightarrow \neg Oxy)$$

If we add (P2) and (UF) on top of the said axiom, the resultant theory is equivalent to (P2) + (SSP) + (UF). However, by the definition of P, it can be easily shown that (P1) and (P2) are logical truths, that is, they must hold. Hence the aforementioned theory is equivalent to **GEM**. If UF is replaced by Postulate II (this is what has been done by Leonard and Goodman), the resultant theory is equivalent to Leśniewski's theory.

The second is to take a binary predicate O, whose intended meaning is “overlapping,” as the primitive (Goodman 1951; Niebergall 2014). Then P is defined as follows.

$$Pxy =_{\text{df}} \forall z (Ozx \rightarrow Ozy)$$

This definition is as intuitive as the previous one. Under this definition, (P1), (P3), and (SSP) are logical truths. Besides, equality is defined as follows (Goodman 1951).

$$x = y =_{\text{df}} \forall z (Ozx \leftrightarrow Ozy)$$

Under this definition and the definition of P, (P2) is a logical truth. Furthermore, the following is an axiom (Goodman 1951).

$$\text{Oxy} \leftrightarrow \exists z \forall w (\text{Owz} \rightarrow (\text{Owx} \wedge \text{Owy}))$$

This is equivalent to $\text{Oxy} \leftrightarrow \exists z (\text{Pzx} \wedge \text{Pzy})$, and this is exactly the definition of O when we take P as the primitive. Therefore, what we have now is a theory equivalent to **EM**, and we will get a theory equivalent to **GEM** if (UF) is added (this is almost what has been done by Goodman 1951).

The third is to use a binary predicate C, whose intended meaning is “being connected to” or “contact” (Whitehead 1929; Clarke 1981; Casati and Varzi 1999; Pratt-Hartmann 2007). Then P is defined as follows.

$$\text{Pxy} =_{\text{df}} \forall z (\text{Czx} \rightarrow \text{Czy})$$

This definition still looks intuitive. The following are two basic axioms of C. They say that the relation “being connected to” is reflexive and symmetric, which should be the case.

$$(C1) \forall x \text{Cxx}$$

$$(C2) \forall x \forall y (\text{Cxy} \rightarrow \text{Cyx})$$

However, this time we will get a new theory, for there seems no way to define “being connected to” in terms of “being a part of.” The previous two alternatives cannot get anything new since “overlapping” and “being a part of” can define each other. It has been pointed out that “being connected to” is a topological concept which cannot be reduced to a mereological one (Casati and Varzi 1999). The indefinability here is not a mathematical proposition that has been proved. Instead, it is supported by philosophical arguments. For instance, “being connected to” at least involves our intuitions about the relations between objects which take up spaces, and then one might turn to a mathematical model of space to argue the said indefinability. A concrete example is to take as the domain the set of nonempty regular open sets in $\mathbb{R}^2$ with the standard topology and interpret P as the set inclusion. Call the

foregoing structure M, and it can be shown that M is a model of **GEM**. Then the “intended” interpretation of C is arguably as follows: x is connected to y if the closure of x intersects the closure of y. But it is not difficult to find an automorphism on M such that connectedness is not preserved (also see the model defined below).

The new theory which will be generated by this alternative is called mereotopology. By the definition of P in terms of C, (P1) and (P3) are logical truths. Since intuitively (P2) should be the case, we will set it as an axiom, but then there seems no clue to what else axioms concerning P should be posited. Instead of trying to get axioms based on some intuitive reading on the meaning of “being connected to,” a different approach is to set an “intended” model or a class of intended models first and then look into what axioms hold there. For instance, one might take the set RO of regular open sets of $\mathbb{R}^2$ and interpret C as $\{(x, y) \in \text{RO} \times \text{RO} : \text{c}(x) \cap \text{c}(y) \neq \emptyset\}$, where c is the closure operator induced by the standard topology on $\mathbb{R}^2$, then it can be shown that P will be the set inclusion, and therefore every desired axiom concerning P holds in that model. Let’s compare this model with M defined above. The latter shows that if the domain is RO and the interpretation of P is the intended one, the intended interpretation of C cannot be preserved no matter how C is defined by using P. But the former shows that if the domain is RO and the interpretation of C is the intended one, then the definition of P specified above does preserve the intended interpretation of P.

Some authors suggest that both P and C should be kept as primitives (Casati and Varzi 1999), and then theories can be formed by axioms from both sides. For instance, **GEM** + (C1) + (C2) + (C3): $\text{Pxy} \rightarrow \forall z (\text{Czx} \rightarrow \text{Czy})$ is called “general extensional mereotopology” (**GEMT**), which is much stronger than **GEM**. Mereotopology as a proper extension of mereology employs an additional topological notion which actually goes beyond the intended content of this entry, and hence we won’t pursue it further.

Models of Mereological Theories

A model of a theory is a structure (of the formal language of the theory considered) which satisfies that theory (“theory” here means “logical theory,” same below). Since any mereological theory must contain **GM**, any of its models must be a partial ordering. Let $\text{Mod}(\mathbf{T})$ stand for the class of the models of **T**. Then $R \in \text{Mod}(\mathbf{GM})$ if and only if R is a partial ordering. Suppose henceforth that any relation R considered is a partial ordering. Let $\text{Dom}(R)$ be the domain of R and the ordering in R be $<_R$ (we will assume that $\text{Dom}(R)$ is non-empty and that $<_R$ is a strict ordering). For any member x in R , let $\text{Initial}_R(x) = \{y \in \text{Dom}(R) : y <_R x\}$ and $\text{Initial}^*_R(x) = \{x\} \cup \text{Initial}_R(x)$. Then we have the following facts (we will omit the subscript if the omission causes no confusion).

$R \in \text{Mod}(\mathbf{GM}+(\mathbf{EP}))$ if and only if for any a, b in $\text{Dom}(R)$, if $a \neq b$, then $\text{Initial}(a) \neq \text{Initial}(b)$ or $\text{Initial}(a) = \text{Initial}(b) = \emptyset$.

$R \in \text{Mod}(\mathbf{MM})$ if and only if for any a, b in $\text{Dom}(R)$, if $a < b$, then for some $c < b$, $\text{Initial}^*(c) \cap \text{Initial}^*(a) = \emptyset$.

$R \in \text{Mod}(\mathbf{EM})$ if and only if for any a, b in $\text{Dom}(R)$, if not $a \leq b$, then for some $c \leq a$, $\text{Initial}^*(c) \cap \text{Initial}^*(b) = \emptyset$.

$R \in \text{Mod}(\mathbf{MEM})$ if and only if $R \in \text{Mod}(\mathbf{MM})$ and for any a, b in $\text{Dom}(R)$, if $\text{Initial}^*(a) \cap \text{Initial}^*(b) \neq \emptyset$, then $\text{Initial}^*(a) \cap \text{Initial}^*(b)$ has a greatest member (i.e., there is some $c \in \text{Initial}^*(a) \cap \text{Initial}^*(b)$ such that for all $d \in \text{Initial}^*(a) \cap \text{Initial}^*(b)$, either $c = d$ or $d < c$).

$R \in \text{Mod}(\mathbf{CM})$ if and only if for any a, b in $\text{Dom}(R)$, if $\text{Initial}^*(a) \cap \text{Initial}^*(b) \neq \emptyset$, then $\text{Initial}^*(a) \cap \text{Initial}^*(b)$ has a greatest member, and for any a, b, c in $\text{Dom}(R)$, if $a \leq c$ and $b \leq c$, then there is some d in $\text{Dom}(R)$ such that for any $e \in \text{Initial}^*(d)$, $\text{Initial}^*(e) \cap (\text{Initial}^*(a) \cup \text{Initial}^*(b)) \neq \emptyset$ and that for any $e \in \text{Initial}^*(a) \cup \text{Initial}^*(b)$, $\text{Initial}^*(e) \cap \text{Initial}^*(d) \neq \emptyset$.

$R \in \text{Mod}(\mathbf{CEM})$ if and only if $R \in \text{Mod}(\mathbf{MEM})$, and for any a in $\text{Dom}(R)$, $\text{Initial}^*(a)$ is an upper semilattice with a greatest member (obviously, the greatest member of $\text{Initial}^*(a)$ is a), where for any $b, c \in \text{Initial}^*(a)$, if $b \leq c$ or $c \leq b$, their join is $\max\{b, c\}$; otherwise, d is the join of b and c if and only if $\text{Initial}^*(b) \cup \text{Initial}^*(c) \subseteq \text{Initial}^*(d)$ and for any

$e \in \text{Initial}^*(d)$, $\text{Initial}^*(e) \cap (\text{Initial}^*(b) \cup \text{Initial}^*(c)) \neq \emptyset$ (An upper semilattice is a structure of the form $(A, \vee)$, where A is a nonempty set and $\vee$ is a binary function on A such that $\vee$ is commutative, associative, and idempotent (i.e., $x \vee x = x$). By defining $x \leq y$ if and only $x \vee y = y$, we can get a partial ordering $(A, \leq)$ from $(A, \vee)$. For any upper semilattice $(A, \vee)$, if there is an x in A such that for any y in A , $x \vee y = x$, $(A, \vee)$ is called an upper semilattice with a greatest member, which must be unique.).

$R \in \text{Mod}(\mathbf{CEM}+(\mathbf{G}))$ if and only if $R \in \text{Mod}(\mathbf{MEM})$ and R is an upper semilattice with a greatest member, where for any $b, c \in \text{Dom}(R)$, if $b \leq c$ or $c \leq b$, their join is $\max\{b, c\}$; otherwise, d is the join of b and c if and only if $\text{Initial}^*(b) \cup \text{Initial}^*(c) \subseteq \text{Initial}^*(d)$ and for any $e \in \text{Initial}^*(d)$, $\text{Initial}^*(e) \cap (\text{Initial}^*(b) \cup \text{Initial}^*(c)) \neq \emptyset$.

$R \in \text{Mod}(\mathbf{CEM} + (\mathbf{C}))$ if and only if (i) R is a Boolean algebra with 0 removed or (ii) R is a Boolean algebra with both 0 and 1 removed (again, $\text{Dom}(R)$ cannot be empty).

$R \in \text{Mod}(\mathbf{CEM} + (\mathbf{C}) + (\mathbf{G}))$ if and only if R is a Boolean algebra with 0 removed.

$R \in \text{Mod}(\mathbf{GEM})$ if and only if R is a Boolean algebra with 0 removed in which every definable subset of the domain has a join (Consider a structure M of a formal language L . A subset S of $\text{Dom}(M)$ is definable if and only if there is a formula $\alpha(x, y_1, \dots, y_n)$, for some $n \geq 0$ (if $n = 0$, such a formula is $\alpha(x)$) such that $S = \{a \in \text{Dom}(M) : M \models \alpha(x, y_1, \dots, y_n)[a, b_1, \dots, b_n]\}$, where $b_1, \dots, b_n$ are n fixed members in $\text{Dom}(M)$. Those n fixed members are called “parameters.” Hence, what we called “definable” here is sometimes called “definable with parameters.”).

It is notable that any mereological theory at least as strong as **EM** has the following theorem: $\forall z(\text{Oxz} \rightarrow \text{Oyz}) \rightarrow \text{Pxy}$, and hence the finite sum of two underlapping members must be unique if it exists (the finite product, if it exists, must be unique because of anti-symmetry).

Now, as an example, let’s explain briefly why $R \in \text{Mod}(\mathbf{CEM} + (\mathbf{C}) + (\mathbf{G}))$ if and only if R is a Boolean algebra with 0 removed. Consider a Boolean algebra B with 0 removed. By Stone Representation Theorem, B can be thought of as a field of sets with the empty set removed (A field of sets is of the form $(A, \cup, \cap, \setminus, \emptyset, X)$, where X is

a nonempty set and $A \subseteq P(X)$ is a nonempty set of some subsets of X such that A is closed under union, intersection, and complementation with respect to X , that is, for any $x, y \in A$, $x \cup y \in A$, $x \cap y \in A$, and $X \setminus x = \{a \in X: a \notin x\} \in A$.). Let the interpretation of Pxy be $x \cup y = y$, that is, $x \subseteq y$. B satisfies **GM** since $\subseteq$ is a partial ordering. B obviously satisfies (C) and (G). B satisfies (SSP) since if not $x \subseteq y$, then $x \setminus y = \{z \in x: z \notin y\}$ cannot be empty (note that $x \setminus y$ equals to the intersection of x and the complement of y). B satisfies (FS) since the union of two members can serve as their finite sum. B satisfies (FP) since the intersection of two members can serve as their finite product (if the intersection of two members is empty, then they do not overlap since $\emptyset$ has been removed). Conversely, consider any model M which satisfies **CEM** + (C) + (G). Since finite sum can serve as join and finite product can serve as meet, it should be clear that M can be converted in a straightforward way into a Boolean algebra with 0 removed.

In general, the class of mereological structures can be classified into the following five mutually disjoint subclasses: finite structures (each of which must be atomic), atomic infinite structures, atomless structures (each of which must be infinite), mixed structures each of which has atomless members as well as finitely many atoms, and, finally, mixed structures each of which has atomless members as well as infinitely many atoms. Of course, this also means that for any mereological theory T , all models of T can be classified into the said five mutually disjoint classes (although some of them might be empty).

In order to clarify some further issues, we need the following definitions.

For any formal language L , any class K of L -structures, let $\text{Th}(K) = \{\alpha \in \text{Formula}(L): \alpha \text{ is a sentence and for any } M \in K, M \models \alpha\}$, that is, $\text{Th}(K)$ is the set of all sentences each of which is satisfied by all structures in K ($\text{Th}(K)$ is called the theory of K). $\text{Th}(K)$ is axiomatizable if and only if there is a set S of some sentences in L such that $\text{Mod}(S) = K$. Note that if K is empty, $\text{Th}(K)$ will be the set of all sentences in L .

Some notable facts are as follows. It is possible that $\text{Th}(K)$ is not axiomatizable. For example, it is well-known that for any first-order language L ,

the theory of the class K of all finite L -structures is not axiomatizable (By the compactness theorem (which only applies to the first-order logic), a sentence which is satisfied by every finite structure will also be satisfied by an infinite structure. See Enderton (2001) for the compactness theorem and its applications.). It is also possible that $\text{Th}(K)$ is axiomatizable but not recursively axiomatizable. For instance, consider the first-order language of arithmetic and the standard arithmetic structure M . Obviously, $\text{Th}(\{M\})$ is axiomatizable since it is exactly the set of sentences which are satisfied by M , but it is a well-known result that $\text{Th}(\{M\})$ is not recursively axiomatizable (The simplified first-order language of arithmetic uses the equality symbol and has only two binary function symbols $+$ and $\cdot$. The standard model of arithmetic has the set of natural numbers as its domain, and the interpretations of those function symbols are exactly those which we are familiar with. It follows from Gödel's incompleteness theorem that the first-order theory of the standard arithmetic model is not recursively axiomatizable. See Chap. 3 in Enderton (2001) for expositions concerning this issue.). Recall that Leśniewski's theory is not first-order axiomatizable. By the notation defined here, that means $\text{Th}(K)$ is not axiomatizable, where K is the class of complete Boolean algebras with 0 removed and where the language considered is the first-order language of mereology. As a matter of fact, $\text{Th}(K) = \mathbf{GEM}$, but $K \subset \text{Mod}(\mathbf{GEM})$ since some models of **GEM** are not complete Boolean algebras with 0 removed.

Let $(n \text{ members})$ stand for an axiom which says that there are exactly n members.

Let $(\text{Infinite members})$ stand for a list of infinitely many axioms α_n , for any natural number $n \geq 2$, where each α_n says that there are at least n members.

Let $(n \text{ atoms})$ stand for an axiom which says that there are exactly n atoms.

Let (Infinite atoms) stand for a list of infinitely many axioms α_n , for any natural number $n \geq 1$, where each α_n says that there are at least n atoms.

It should be clear that each of the axioms expressed in English above can be translated into a first-order sentence.

Consider models of $\mathbf{CEM} + (C) + (G)$. It is easy to show that any finite model of $\mathbf{CEM} + (C) + (G)$ will be a complete Boolean algebra with 0 removed and therefore will also satisfy $\mathbf{GEM}$. Moreover, each finite model of $\mathbf{CEM} + (C) + (G)$ has exactly $2^k - 1$ members, for some $k \geq 1$ (because the domain of a finite Boolean algebra must have exactly 2^k members, for some $k \geq 1$). Since any two finite Boolean algebras with the same cardinality are isomorphic, the theory of the class of models each of which has exactly $2^k - 1$ members, for some $k \geq 1$, is recursively axiomatized by $\mathbf{CEM} + (C) + (G) + (2^k - 1 \text{ members})$. The foregoing facts are easy to see. Some more interesting results can be shown about infinite models.

Any two atomic infinite models of $\mathbf{CEM} + (C) + (G)$ are elementarily equivalent (Tsai 2013a). It follows that $\mathbf{CEM} + (C) + (G) + (A) + (\text{Infinite members})$ is a complete theory. Let K_1 be the class of atomic infinite models of $\mathbf{CEM} + (C) + (G)$. Then $\text{Th}(K_1) = \mathbf{CEM} + (C) + (G) + (A) + (\text{Infinite members})$. Furthermore, since $\mathbf{GEM} + (A) + (\text{Infinite members})$ is a consistent theory, the foregoing result implies that $\text{Th}(K_1) = \mathbf{GEM} + (A) + (\text{Infinite members})$, and therefore, $\mathbf{GEM}$ and $\mathbf{CEM} + (C) + (G)$ share the same atomic infinite models.

Any two atomless models of $\mathbf{CEM} + (C) + (G)$ are elementarily equivalent (they are isomorphic if both are countable) (Tsai 2013a). It follows that $\mathbf{CEM} + (C) + (G) + (AL)$ is a complete theory. Let K_2 be the class of atomless models of $\mathbf{CEM} + (C) + (G)$. Then $\text{Th}(K_2) = \mathbf{CEM} + (C) + (G) + (AL)$. The said result implies that $\mathbf{CEM} + (C) + (G)$ and $\mathbf{GEM}$ share the same atomless models.

Any two mixed models of $\mathbf{CEM} + (C) + (G)$ which have the same number of finitely many atoms are elementarily equivalent (they are isomorphic if both are countable) (Tsai 2018). It follows that for each $n \geq 1$, $\mathbf{CEM} + (C) + (G) + \exists x \text{Gunk}(x) + (n \text{ atoms})$ is complete. For each $n \geq 1$, let K_{3n} be the class of mixed models of $\mathbf{CEM} + (C) + (G)$ each of which has exactly n atoms. Then $\text{Th}(K_{3n}) = \mathbf{CEM} + (C) + (G) + \exists x \text{Gunk}(x) + (n \text{ atoms})$. The foregoing

result implies that $\mathbf{CEM} + (C) + (G)$ and $\mathbf{GEM}$ share the same mixed models each of which has finitely many atoms.

However, we can find a mixed structure with infinitely many atoms which satisfies $\mathbf{CEM} + (C) + (G)$ but does not satisfy $\mathbf{GEM}$ (Tsai 2018). This shows that $\mathbf{CEM} + (C) + (G) + \exists x \text{Gunk}(x) + (\text{Infinite atoms})$ is incomplete. The said mixed model does not satisfy $\mathbf{GEM}$ because its domain does not have a member which is composed of exactly all atoms, which contradicts (UF). Let K_4 be the class of mixed models of $\mathbf{CEM} + (C) + (G) + \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Atom}(x) \wedge \text{Oyx}))$ each of which has infinitely many atoms. It turns out that any two models in K_4 are elementarily equivalent, which implies that $\mathbf{CEM} + (C) + (G) + \exists x \text{Gunk}(x) + (\text{Infinite atoms}) + \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Atom}(x) \wedge \text{Oyx}))$ is a complete theory and is a recursive axiomatization of $\text{Th}(K_4)$.

It follows from the foregoing analyses of models that $\mathbf{GEM} = \mathbf{CEM} + (C) + (G) + \exists x \text{Atom}(x) \rightarrow \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Atom}(x) \wedge \text{Oyx}))$. Moreover, by (C), if the fusion of all atoms exists, then the fusion of all gunks exists too, for they complement each other. That is, under $\mathbf{CEM} + (C) + (G)$, $\exists x \text{Atom}(x) \rightarrow \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Atom}(x) \wedge \text{Oyx}))$ is equivalent to $\exists x \text{Gunk}(x) \rightarrow \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Gunk}(x) \wedge \text{Oyx}))$, and therefore, we also have $\mathbf{GEM} = \mathbf{CEM} + (C) + (G) + \exists x \text{Gunk}(x) \rightarrow \exists z \forall y (\text{Oyz} \leftrightarrow \exists x (\text{Gunk}(x) \wedge \text{Oyx}))$.

Decidability Concerning First-Order Mereological Theories

We are only concerned with first-order theories in this section. A logical theory T is called “decidable” if and only if there is an effective procedure such that for any sentence α in the language of T , it can be answered by using such a procedure whether or not α is a theorem of T . Of course, a theory is undecidable if it is not decidable.

The definition of decidability above resorts to the notion of “effective procedure.” But what is an

effective procedure? As characterized in Enderton (2001), there must be finitely many exact instructions explaining how to execute the procedure and these instructions “must not demand any brilliance or originality on the part of the person or machine following them,” and the procedure “must avoid random devices” and must produce an answer, “yes” or “no,” after a finite number of steps. But such a characterization is apparently informal, and therefore the definition of decidability above is also informal.

However, the symbols of a “reasonable” first-order language can be mapped into natural numbers in such a way that we can check within finitely many steps whether or not a natural number is the code of a symbol under that mapping. Then each term t , which is a finite sequence of symbols, can also be coded into a natural number, and furthermore, each formula α , which again is a finite sequence of symbols, can be coded into a natural number too, and such a number is called the Gödel number of α (see Sect. 3.4 in Enderton (2001) for a way of coding). In this light, a set of formulas in a first-order language is decidable if and only if the set of Gödel numbers of those formulas is decidable (i.e., there is an effective procedure to check whether or not a natural number is in that set). But by Church’s thesis (see Enderton 2001), the notion of “decidable” and the notion of “recursive” share the same extension, that is, a set of natural numbers is decidable if and only if it is recursive (i.e., its characteristic function is recursive). But “recursive” is a notion which has a rigorous mathematical definition. Since Church’s thesis, which strictly speaking is not a mathematical proposition, has been widely accepted (no counterexample has been found so far, and there are pieces of evidence which support it), in the following “decidable” will be taken as equivalent to “recursive.”

A set of natural numbers is effectively enumerable if there is an effective procedure to enumerate its members one by one. This property is weaker than decidability since it does not require that the effective procedure can also effectively enumerate those nonmembers one by one (for the difference between “decidable” and “effectively enumerable,” see Enderton 2001). Again, by Church’s

thesis, we take “effectively enumerable” as equivalent to “recursively enumerable.”

There is a negative property which is stronger than undecidability called “finite inseparability” which is defined as follows (see Monk (1976); the version presented here has been rephrased by the present writer).

Two sets A and B of natural numbers are *effectively inseparable* if and only if $A \cap B = \emptyset$ and there is an effective procedure via which, given any two recursively enumerable sets C and D such that $A \subseteq C$, $B \subseteq D$, and $C \cap D = \emptyset$, we can find a natural number which does not belong to $C \cup D$.

A theory T based on a language L is *finitely inseparable* if and only if $\{\#\alpha: \alpha \text{ is a sentence in } L \text{ and is satisfied by every structure of } L\}$ and $\{\#\alpha: \alpha \text{ is a sentence in } L \text{ whose negation is satisfied by some finite model of } T\}$ are effectively inseparable, where $\#\alpha$ stands for the Gödel number of α .

By the definition of finite inseparability, it is easy to see that any finitely inseparable theory must be undecidable.

It is known that **GM** is finitely inseparable (see Monk 1976). We also have the following results (Tsai 2009, 2011, 2013a, b). For any first-order mereological theory T , T is finitely inseparable if **GM** $\subset$ $T \subseteq$ **CEM**. **CEM** + (G) is not finitely inseparable (since it shares finite models with **CEM** + C), but is undecidable. **CEM** + (C) is decidable, and hence any of its finite extensions (for any theory T , a finite extension of T is a theory axiomatized by T plus some finitely many axioms), such as **CEM** + (C) + (G), is decidable too. Furthermore, as mentioned earlier, **GEM** is a finite extension of **CEM** + (C) + (G) and therefore is also decidable.

Also observe that each of the following mereological theories is consistent and complete (see section “Models of Mereological Theories” above; also see Tsai (2018) for proofs of consistency).

For each natural number $k \geq 1$, **GEM** + ($2^k - 1$ members).

GEM + (A) + (Infinite members)

GEM + (AL)

For each natural number $k \geq 1$, **GEM** + $\exists x \text{Gunk}(x)$ + (k atoms).

GEM + $\exists x \text{Gunk}(x)$ + (Infinite atoms)

Moreover, any first-order mereological theory T can be represented in the following way. Let $T_1 = \cap \{\text{Th}(K_n): K_n \text{ is the class of models of } T \text{ each of which has exactly } n \text{ members, for some } n \geq 1\}$, $T_2 = \text{Th}(K_{AI})$, where K_{AI} is the class of atomic infinite models of T , $T_3 = \text{Th}(K_{AL})$, where K_{AL} is the class of atomless models of T , $T_4 = \cap \{\text{Th}(K_{mn}): K_{mn} \text{ is the class of mixed models of } T \text{ each of which has exactly } n \text{ atoms, for some } n \geq 1\}$, $T_5 = \text{Th}(K_{mi})$, where K_{mi} is the class of mixed models of T each of which has infinite atoms. Then it must be the case that $T = T_1 \cap T_2 \cap T_3 \cap T_4 \cap T_5$. When T is at least as strong as **GEM**, that is, $\text{GEM} \subseteq T$, those K_n whose subscript is not of the form $2^k - 1$ can be dropped, and each $\text{Th}(K_n)$ is equal either to the set of all sentences (which means that $T + (n \text{ members})$ is inconsistent) or to $\text{GEM} + (n \text{ members})$, each $\text{Th}(K_{mn})$ is equal either to the set of all sentences (which means that $T + \exists x \text{Gunk}(x) + (n \text{ atoms})$ is inconsistent) or to $\text{GEM} + \exists x \text{Gunk}(x) + (n \text{ atoms})$, $\text{Th}(K_{AI})$ is equal either to the set of all sentences (which means that $T + (A) + (\text{Infinite members})$ is inconsistent) or to $\text{GEM} + (A) + (\text{Infinite members})$, $\text{Th}(K_{AL})$ is equal either to the set of all sentences (which means that $T + (AL)$ is inconsistent) or to $\text{GEM} + (AL)$, and $\text{Th}(K_{mi})$ is equal either to the set of all sentences (which means that $T + \exists x \text{Gunk}(x) + (\text{Infinite atoms})$ is inconsistent) or to $\text{GEM} + \exists x \text{Gunk}(x) + (\text{Infinite atoms})$.

So if $\text{GEM} \subseteq T$, each of T_2 , T_3 , and T_5 is recursively axiomatizable and complete, and each $\text{Th}(K_n)$ or $\text{Th}(K_{mn})$, for $n \geq 1$, is also recursively axiomatizable and complete (an inconsistent theory will contain all sentences and hence by definition must be recursively axiomatizable and complete). But a recursively axiomatizable complete first-order theory must be decidable (see Enderton 2001). Therefore, T is the intersection of enumerably many decidable theories, each of which is a recursively axiomatizable complete extension of T (it is obvious that the aforementioned theories can be enumerated one by one). If each of the said extensions of T is consistent, T will be equal (logically equivalent) to **GEM**, which means that some of the extensions must be

inconsistent if T is strictly stronger than **GEM**. Then by Ershov's theorem (Monk 1976), for any T such that $\text{GEM} \subseteq T$, T will be decidable if and only if there is an effective procedure to check whether an extension of T in the said enumeration is consistent or not.

Consider the following axiom which is called "local complementation" (LC) (Tsai 2013b, 2015).

$$(LC) \forall x \forall y ((Ppxy \wedge \neg \forall z Pzy) \rightarrow \exists z (PPzy \wedge \neg Ozx \wedge \forall w (Owy \leftrightarrow (Owx \vee Owz))))$$

CEM + (LC) is still undecidable and in fact finitely inseparable (see the proof of Theorem 1 in Tsai (2009)). It is not difficult to check that the model defined in that proof also satisfies **CEM** + (LC). Moreover, it is known that any first-order mereological theory which is recursively axiomatizable by using axioms found in the literature (see section "First-Order Mereological Axioms and Theories") and which has at least an atomic model is undecidable if it does not imply (LC) (Tsai 2015). **CEM** + (C) can prove (LC) and in fact $\text{CEM} + (LC) \subset \text{CEM} + (C)$. An interesting question is whether we can find a decidable theory T such that $\text{CEM} + (LC) \subset T \subset \text{CEM} + (C)$. The answer is positive. For instance, let T be the intersection of $\text{CEM} + (C)$ and $\text{CEM} + (LC) + (A) + (G) + (\text{Infinite members}) + \forall x \neg \exists z \forall w (Pwx \leftrightarrow \neg Owz)$. Since these two theories are decidable (Tsai 2013b), T is decidable too. It is trivial that $\text{CEM} + (LC) \subseteq T$, but they cannot be identical since T is decidable, while **CEM** + (LC) is not.

Fuzzification of Mereology

Parthood relation might come in degree. Suppose degrees range over real numbers from 0 to 1. Consider a mereological structure. Let π be a binary function from the Cartesian product of its domain to $[0, 1]$. As a preliminary example, it seems reasonable to suggest the following three principles (Varzi 2016).

$$(\pi 1) \pi(x, x) = 1$$

$$(\pi 2) \pi(x, z) \geq \min(\pi(x, y), \pi(y, z))$$

$$(\pi 3) \text{ If } \pi(x, y) = 1 \text{ and } \pi(y, x) = 1, \text{ then } x = y.$$

The foregoing example gives us a basic idea of how to devise a semantics for “fuzzy” mereology. However, to get to a logical theory, an appropriate syntactics, that is, a well-defined formal language of “fuzzy” mereology, is in need. The following is a first-order language which can do the job (the untyped language to be defined is a rephrased version based on the typed language given in Polkowski and Skowron 1996). On top of the first-order logical symbols (including equality), the language has two monadic predicates *Object*(*x*) and *Degree*(*x*), one binary predicate *<*(*x*, *y*) and one ternary predicate *P*(*x*, *y*, *z*). The intended meanings of those symbols are as follows. *Object*(*x*) means that *x* is an object, *Degree*(*x*) means that *x* is a degree, *<*(*x*, *y*) means that degree *x* < degree *y*, and *P*(*x*, *y*, *z*) means that the degree of *x* being a part of *y* is *z*. The following are some basic axioms.

1. $\forall x \forall y ((\text{Object}(x) \wedge \text{Object}(y)) \rightarrow \exists! z (\text{Degree}(z) \wedge P(x, y, z)))$
2. $\forall x (\text{Degree}(x) \rightarrow \neg <(x, x))$
3. $\forall x \forall y \forall z ((\text{Degree}(x) \wedge \text{Degree}(y) \wedge \text{Degree}(z) \wedge <(x, y) \wedge <(y, z)) \rightarrow <(x, z))$
4. $\forall x \forall y ((\text{Degree}(x) \wedge \text{Degree}(y)) \rightarrow (<(x, y) \vee <(y, x) \vee x = y))$
5. $\exists x (\text{Degree}(x) \wedge \forall y (\text{Degree}(y) \rightarrow \neg <(y, x)))$
6. $\exists x (\text{Degree}(x) \wedge \forall y (\text{Degree}(y) \rightarrow \neg <(x, y)))$

Note that (2)–(6) mean that degrees are linear ordered by *<* with a greatest degree and a least degree. Since either is unique, we can use 1 to denote the greatest degree and 0 to denote the least degree, that is, 1 and 0 are two definable constants. Moreover, for any two degrees *x*, *y*, *min*(*x*, *y*) is definable, and for any two objects *x*, *y*, *π*(*x*, *y*) is definable too (Let’s briefly explain how to define those symbols. The formula *x* = 0 can be defined

as an abbreviation of $\text{Degree}(x) \wedge \forall y (\text{Degree}(y) \rightarrow \neg <(y, x))$. The formula *x* = 1 can be defined as an abbreviation of $\text{Degree}(x) \wedge \forall y (\text{Degree}(y) \rightarrow \neg <(x, y))$. The formula *π*(*x*, *y*) = *z* can be defined as an abbreviation of $\text{Object}(x) \wedge \text{Object}(y) \wedge \text{Degree}(z) \wedge P(x, y, z)$. Finally, the formula *min*(*x*, *y*) = *z* can be defined as an abbreviation of $\text{Degree}(x) \wedge \text{Degree}(y) \wedge ((<(x, y) \rightarrow x = z) \wedge (\neg <(x, y) \rightarrow y = z))$. With these additional definable symbols, other basic axioms corresponding to the aforementioned (π1)–(π3) can be more succinctly expressed as follows.

7. $\forall x (\text{Object}(x) \rightarrow P(x, x, 1))$
8. $\forall x \forall y \forall z ((\text{Object}(x) \wedge \text{Object}(y) \wedge \text{Object}(z)) \rightarrow \neg <(\pi(x, z), \min(\pi(x, y), \pi(y, z))))$
9. $\forall x \forall y ((\text{Object}(x) \wedge \text{Object}(y) \wedge \pi(x, y) = 1 \wedge \pi(y, x) = 1) \rightarrow x = y)$

The set of objects should be nonempty, and the set of degrees should have at least two members. Moreover, we may also ask that the set of objects be disjoint from the set of degrees and that the union of them be exactly the whole domain. Those requirements can be achieved by the following axioms.

10. $\exists x \text{Object}(x)$
11. $\exists x \exists y (\text{Degree}(x) \wedge \text{Degree}(y) \wedge \neg x = y)$
12. $\forall x (\text{Object}(x) \vee \text{Degree}(x))$
13. $\forall x (\text{Object}(x) \rightarrow \neg \text{Degree}(x))$

Furthermore, since *P*(*x*, *y*, 1) can be thought of as *P*(*x*, *y*) in the non-fuzzy mereology, it should be clear that axioms of non-fuzzy mereology can be translated into the language of fuzzy mereology. Other axioms are still possible, but here only a general layout is intended.

It is easy to define a model which satisfies the aforementioned axioms. In general, any model of them is of the form (U, I), where U is a set which has at least three members and I is an interpretation of the parameters of the language of fuzzy mereology such that each of the following conditions is satisfied.

I(Object) is a nonempty subset of U.

$I(\text{Degree}) = UI(\text{Object})$, and there are at least two members in $I(\text{Degree})$.

$I(<)$ is a bounded linear ordering on $I(\text{Degree})$.

$I(P)$ is a function from $I(\text{Object}) \times I(\text{Object})$ to $I(\text{Degree})$ such that (i) for any $a \in I(\text{Object})$, $I(P)(a, a) = 1$; (ii) for any $a, b \in I(\text{Object})$, $a = b$ if $I(P)(a, b) = 1$ and $I(P)(b, a) = 1$; and (iii) for any $a, b, c \in I(\text{Object})$, $I(P)(a, c) \geq \min(I(P)(a, b), I(P)(b, c))$, where 1 and min are defined in $I(\text{Degree})$ with the ordering $I(<)$.

The foregoing gives a straightforward formalization of fuzzy mereology in which the talk about degrees, which is originally meta-logical, has also been formalized in the object language. As shown above, the domain has two different categories, Object and Degree. If one dislikes this, it is possible to merge two into one, and the idea is that objects can themselves also play the role of degrees. For example, on top of the usual mereological binary predicate P , we can add a binary degree function F and add axioms to put requirements on the range of F , such as $\forall u \forall v ((\exists x \exists y F(x, y) = u \wedge \exists x \exists y F(x, y) = v \wedge \neg P(u, v)) \rightarrow PP(v, u))$. This axiom working with the basic axioms of P suffices to guarantee that the range of F is linear ordered by the proper part relation. Moreover, we may loosen restrictions on degrees. For instance, instead of a linear ordering, degrees may just form a partial ordering. In short, there can be other possible formalizations of fuzzy mereology, and each might be motivated by an intended application that the author has in his or her mind (For instance, the theory named “rough mereology” in Polkowski and Skowron (1996) is intended to give, as the title of the paper suggests, “a new paradigm for approximate reasoning.” The language of rough mereology has no equality symbol, and for the authors’ purpose, besides the basic axioms (A1)~(A3) which intuitively should hold (they roughly correspond to $(\pi 1) \sim (\pi 3)$), rough mereology has an axiom (A4) which says that there is a minimal member and an axiom (A5) which says the same thing as (SSP) and an axiom schema (A6)_n which says that for any $n \geq 2$, every finite subset consisting of some n members of the domain has a fusion.).

One more thing is that the theory (let’s call it ground fuzzy mereology, **GFM**) axiomatized by

the 13 axioms listed above can be shown to be finitely inseparable and hence is undecidable. The idea of proof is to show that any finite partial ordering can be extended to a finite model of **GFM** in such a way that the original partial ordering can be defined in that model. Consider a finite partial ordering (A, R) , where R is a strict ordering on A . Extend it to another finite partial ordering (A', R') , where R' is a strict ordering on A' , in the following two steps. Step One: For each R -minimal member x in A , add a new member y , and define $yR'x$ as well as define $yR'z$, for any z in A such that xRz . Step Two: Add two isolated members 0 and 1 to the partial ordering generated after doing Step One. That is to say, $(A, R) \subseteq (A', R')$ and A' has exactly two members 0 and 1 such that for any x in A' , not $xR'0$, not $0R'x$, not $xR'1$, and not $1R'x$ (this is why 0 and 1 are called “isolated members”). Then (A', R') can be turned to a finite model M of **GFM** in the following way. $\text{Dom}(M)$, the domain of M , is A' , $I(\text{Degree}) = \{0, 1\}$, $I(\text{Object}) = A' \setminus \{0, 1\}$, $I(<) = \{(0, 1)\}$, which means $0 < 1$, and $I(P) = \{(x, x, 1): x \in I(\text{Object})\} \cup \{(x, y, 1): x \in I(\text{Object}), y \in I(\text{Object}) \text{ and } xR'y \text{ in } (A', R')\} \cup \{(x, y, 0): x \in I(\text{Object}), y \in I(\text{Object}), x \neq y \text{ and not } xR'y \text{ in } (A', R')\}$. It is easy to see that the model M thus defined satisfies **GFM**. Finally, the original (A, R) can be recovered by $I(\forall) = \exists y(y \neq x \wedge P(y, x, 1))$ and $I(R) = x \neq y \wedge P(x, y, 1)$, that is, $A = \{a \in \text{Dom}(M): M \models \exists y(y \neq x \wedge P(y, x, 1)[a])\}$ and $R = \{(a, b) \subseteq \text{Dom}(M) \times \text{Dom}(M): M \models x \neq y \wedge P(x, y, 1)[a, b]\}$. Then since the theory of partial orderings is finitely inseparable, **GFM** is finitely inseparable too (This is owing to a general theorem about finite inseparability; see Monk (1976). Also see Tsai (2009, 2011) for expositions and applications of the said general theorem.).

Future Directions

Classical mereology as a second-order theory is something like the theory of complete Boolean algebras, and its first-order approximation, that is, general extensional mereology, is something like the elementary theory of Boolean algebras extended by adding an axiom schema which

says that every nonempty definable subset of the domain has a join. It was certainly not the intention of early authors of mereology to formulate theories in such a way that they are almost equivalent to those of Boolean algebras. This is just an accidental fact discovered later. Despite the said similarity, mereological theories can still serve as a useful tool for the reasoning involving the notion of “part,” for the formal language employs directly predicates concerning such a notion. But this is not the case for theories of Boolean algebras, in whose formal language every statement is a mathematical equation. Besides, many mereological axioms as well as mereological theories are philosophically motivated (Simons 1987; Varzi 2016). It seems fair to say that those theories reveal some important aspects about how we human beings carry out reasoning about the notion “part.” But this is hardly a concern of the algebraic approach.

There are still some issues concerning the notion “part” with which those mereological theories formulated so far cannot deal properly. As we can see in the literature, there are philosophical concerns about changing parts without changing the identity, for example, normally we do not think that changing a tire of a car would change the identity of that car. In order to account for that kind of issues, an imaginable and seeming natural proposal is to differentiate parts which are necessary for the identity of an object from those which are not. This would call for a theory something like “modal mereology,” for modal logic is traditionally known as a formal theory about “necessity.” There are some attempts which can be found in the literature, for instance, Uzquiano (2014), but there is still a lot of room for the development of modal mereology.

Besides, it is a well-known philosophical reflection that a whole is not just its parts “putting together”: those parts should be “organized” in some way so as to form the whole in question, for instance, a heap of parts of a car is not that car. But the traditional approach fails to account for such a reflection, and in fact it seems hopeless if we stick with the binary parthood relation (The problem here is arguably similar to the indefinability of “being connected to” in terms of “being

a part of.” For instance, one might propose that any part within a whole must be self-connected, that is, any two parts of such a part are connected with each other (Whitehead 1920). However, as we have seen, this cannot be done by using the parthood relation alone.). Instead, in order to do justice to the said reflection, we should try to formalize the internal relations between those parts within a whole. Some related attempts, for instance, Fine (1999), can be seen in the literature, but there is still room for further developments. Moreover, arguably, a certain kind of mereotopology can be thought of as a way of formalizing an internal relation of parts within a whole. But this is not the only option, and whether we should subsume mereological notions under topological ones is debatable (As mentioned earlier, Casati and Varzi are against reducing mereological notions to topological ones. However, they propose to keep both predicates, and the resultant theory is too strong in my opinion.). It should be worth looking into what other options there are.

Acknowledgments Part of this work was carried out within research project MOST 107-2410-H-194-089-MY3, kindly funded by the Ministry of Science and Technology of Taiwan.

Bibliography

Primary Literature

- Casati R, Varzi A (1999) *Parts and places*. The MIT Press, Cambridge
- Clarke BL (1981) A Calculus of individuals based on “connection”. *Notre Dame J Form Log* 22:204–218
- Clay RE (1974) Relation of Leśniewski’s mereology to Boolean algebras. *J Symb Log* 39:638–648
- Enderton HB (2001) *A mathematical introduction to logic*, 2nd edn. Harcourt Press, San Diego
- Fine K (1999) Things and their parts. *Midwest Stud Philos* 23:61–74
- Goodman N (1951) *The structure of appearance*. Harvard University Press, Cambridge
- Grzegorczyk A (1955) The Systems of Leśniewski in relation to contemporary logical research. *Stud Logica* 3:77–97
- Kunen K (1980) *Set theory: an introduction to independence proofs*. North-Holland, New York
- Leonard HS, Goodman N (1940) The Calculus of individuals and its uses. *J Symb Log* 5:45–55

- Leśniewski S (1992) *Podstawy ogólnej teorii mnogości. I*. Prace Polskiego Koła Naukowego w Moskwie, Sekcja matematyczno-przyrodnicza, Moskwa, 1916; Eng. trans: Barnett DI: Foundations of the general theory of sets. I. Leśniewski S, In: Surma SJ et al. (eds) Collected works, Vol.1. Kluwer, Dordrecht. pp 129–173
- Monk JD (1976) Mathematical logic. Springer, New York
- Niebergall KG (2014) Mereology. In: Horsten L, Pettigrew R (eds) The Bloomsbury companion to philosophical logic. Bloomsbury Academic, London, pp 271–298
- Pietruszczak A (2015) Classical mereology is not elementarily axiomatizable. *Log Logical Philos* 24(4):485–498
- Pietruszczak A (2018) Metamereologia. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, 2000; English version: Pietruszczak A, Metamereology (trans: Carmody M) The Nicolaus Copernicus University Scientific Publishing House, Toruń
- Polkowski L, Skowron A (1996) Rough mereology: a new paradigm for approximate reasoning. *Int J Approx Reason* 15(4):333–365
- Pratt-Hartmann I (2007) First-order mereotopology. In: Aiello M, Pratt-Hartmann I, van Benthem J (eds) Handbook of spatial logics. Springer, Dordrecht, pp 13–97
- Simons P (1987) Parts: a study in ontology. Clarendon Press, Oxford
- Tarski A (1956a) Foundation of the geometry of solids. In: Logic, semantics, metamathematics (trans: Woodger JH). Clarendon Press, Oxford
- Tarski A (1956b) On the foundations of Boolean algebra. In: Logic, semantics, metamathematics (trans: Woodger JH). Clarendon Press, Oxford
- Tsai H (2009) Decidability of mereological theories. *Log Logical Philos* 18(1):45–63
- Tsai H (2011) More on the decidability of mereological theories. *Log Logical Philos* 20(3):251–265
- Tsai H (2013a) Decidability of general extensional mereology. *Stud Logica* 101(3):619–636
- Tsai H (2013b) A comprehensive picture of the decidability of mereological theories. *Stud Logica* 101(5):987–1012
- Tsai H (2015) Notes on models of first-order mereological theories. *Log Logical Philos* 24(4):469–483
- Tsai H (2018) General extensional mereology is finitely axiomatizable. *Stud Logica* 106(4):809–826
- Uzquiano G (2014) Mereology and modality. In: Kleinschmidt S (ed) Mereology and location. Oxford University Press, Oxford, pp 33–56
- Varzi A (2016) Mereology. In: Stanford encyclopedia of philosophy. <https://plato.stanford.edu/entries/mereology/>
- Whitehead AN (1920) The concept of nature. Cambridge University Press, Cambridge
- Whitehead AN (1929) Process and reality. An essay in cosmology. Macmillan, New York

Books and Reviews

- Baumgartner W (2013) Franz Brentano's mereology. In: Fisette D, Fréchette G (eds) Themes from Brentano. Rodopi, Amsterdam, pp 227–245
- Bittner T, Donnelly M (2007) A temporal mereology for distinguishing between integral objects and portions of stuff. In: Holte R, Howe A (eds) Proceedings of the twenty-second AAAI conference on artificial intelligence. AAAI Press, Vancouver, pp 287–292
- Bleeksmith R, Null G (1991) Matrix representation of Husserl's part-whole-foundation theory. *Notre Dame J Form Log* 32:87–111
- Chellas BF (1980) Modal logic: an introduction. Cambridge University Press, Cambridge
- Chisholm RM (1973) Parts as essential to their wholes. *Rev Metaphys* 26:581–603
- Chisholm RM (1975) Mereological essentialism: further considerations. *Rev Metaphys* 28:477–484
- Clay RE (1981) Leśniewski's mereology. Universidad de Oriente, Cumana
- Fine K (1994) Compounds and aggregates. *Noûs* 28:137–158
- Fitting M, Mendelsohn RL (1999) First-order modal logic. Kluwer Academic Publishers, Dordrecht
- Hodges W (1993) Model theory. Cambridge University Press, Cambridge
- Jech T (2002) Set theory, 3rd edn. Springer, Berlin
- Koppelberg S (1989) Handbook of Boolean algebras, vol 1. North-Holland, Amsterdam
- Lewis DK (1991) Parts of classes. Blackwell, Oxford
- Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer, Dordrecht
- Shoenfield JR (1967) Mathematical logic. Addison-Wesley, London
- Soare R (1987) Recursively enumerable sets and degrees. Perspectives in mathematical logic. Springer, Berlin/New York



Variable Precision Approximations of Rough Sets

Wojciech Ziarko

Department of Computer Science, University of Regina, Regina, SK, Canada

Article Outline

Introduction
Basics of Pawlak's Rough Sets
Variable Precision Rough Set Model
Decision Tables Learned from Data
Dependencies in Probabilistic Decision Tables
Reduction and Significance of Attributes
Significance of Attributes
Final Remarks
Bibliography

Introduction

This entry presents the basics of the variable precision rough set model (VPRSM) (Ziarko 1993, 2001, 2002a, 2005, 2008a, b; Slezak and Ziarko 2003; Beynon 2001; Beynon and Peel 2001; Wei and Zhang 2004; Katzberg and Ziarko 1994). The VPRSM is an extension of the original rough set approach to data analysis and modeling from data, as introduced by Zdzislaw Pawlak (1982, 1991). It aims to take advantage of frequency distribution information when forming approximate definitions of sets and inter-data dependencies. The VPRSM approach was utilized primarily in data mining and machine learning applications. The primary motivation behind research leading to the development of the VPRSM is the imperfections of gathered practical application data. In particular, application data often suffer from the presence of measurement noise, leading to lack of consistency and resulting difficulty to form data

classifications and set approximations, as defined in the Pawlak's rough set model.

The VPRSM is focused on the recognition and modeling of overlap-based relationships between sets, which are most useful when dealing with noisy data. In this approach, the set-overlap relationships are used to construct approximate definitions of otherwise undefinable, or rough, sets (Pawlak 1982, 1991). The main application of this approach is the analysis of data co-occurrence-based dependencies in decision tables derived from data, called probabilistic decision tables in the VPRSM. The probabilistic decision tables are usually learned from data to represent some inter-data item connections, typically for the purpose of data mining or machine learning (see Ziarko 2002b, 2003; Chen and Ziarko 2010). They can also be used as a basis of generalized probabilistic rule extraction algorithms (Ziarko 2002b).

In practical applications of the data-acquired decision tables, one of the key issues is the identification of a minimal subset of attributes, which are discrete functions of measured features, to represent an identified data dependency without any loss, or with minimal loss, of information. The original general idea of attribute reduct, as introduced by Pawlak (1982, 1991), is useful in that respect. As in the original rough set model, the extended notion of reduct defined in the context of VPRSM allows for information-preserving identification of minimal subsets of attributes.

The VPRSM is one among several other extensions and generalizations of Pawlak's rough sets, which inspired significant research interest (see Yao 2007, 2008; Yao and Lin 1996; James 2007; Greco et al. 2002; Yao and Yao 2012; Zhang et al. 2013). In particular, the Bayesian rough set model (Slezak and Ziarko 2003), which is derived from VPRSM, is discussed in more detail in this entry.

This entry is organized as follows. In the next section, we review the fundamentals of Pawlak's rough sets, which include the introduction to approximation spaces and rough set approximations. In section "Variable Precision Rough Set

Model,” the main ideas of the VPRSM are introduced, followed by the presentation of key definitions of the Bayesian rough set model. In section “Decision Tables Learned from Data,” the probabilistic attribute value-based decision tables are presented. The tables represent the classification knowledge with respect to a specific target set. The probabilistic decision tables also represent rough approximations of the target set, as defined in the framework of the VPRSM. In section “Dependencies in Probabilistic Decision Tables,” different kinds of probabilistic dependencies occurring in probabilistic decision tables are introduced. All discussed dependencies are of probabilistic nature and are defined in either the context of variable precision or Bayesian rough set models. They generalize and expand the attribute dependencies introduced by Pawlak in the original rough set theory (Pawlak 1982, 1991). Attribute reduction with respect to introduced dependencies is the subject of section “Reduction and Significance of Attributes.” The demonstrated monotonicity property of the dependencies allows for a definition of the notion of information-preserving minimal subset of attributes called probabilistic reduct. The ability to compute reducts allows for the determination of the importance, or significance, of attributes. This is the subject of section “Significance of Attributes.”

Basics of Pawlak’s Rough Sets

The theory of rough sets grew out of series of seminars organized by Victor Marek on the general subject of mathematical foundations of information systems held at the Department of Mathematics of the University of Warsaw, Poland, in the mid-1970s. It was eventually formalized and first published by Zdzislaw Pawlak (1982) in the early 1980s. The theory touched the basics of representation of uncertain knowledge and reasoning with such knowledge. The precise mathematical definitions, in terms of set theory and logic, of the notions of different kinds of uncertainty, turned out to be quite inspiring to computer scientists working in the areas of data analysis, machine learning, or data mining, while attracting

attention of logicians and mathematicians. The data analysis and machine learning aspects of rough set theory also had important practical applications, as tools for deep analysis of inter-data connections or dependencies, and for the acquisition of generalized rules from data for machine learning (Ziarko 2002b), control, and other applications. A number of software systems were developed to support practical applications of rough set theory, but more work in that area is still required. In what follows, the basics of Pawlak’s rough set theory are presented.

Approximation Space

There are two crucial departure points in the rough set theory.

The first departure point is the availability of a finite set of unique objects of interest, denoted as U and called the *universe*. In a typical practice, the universe corresponds to a set of unique “snapshots” of a dynamic phenomenon, or of some distinct physical objects, for example, individual customers in a market research application.

The second departure point is the existence of a prior ability, or knowledge, to form the classification of the universe of objects of interest into distinct classes. This ability, referred to as *classification knowledge*, is often associated with an external agent, such as a domain expert, for example, who is assumed to *know how* to classify objects into categories. However, in automated systems, such an expert typically is not available. Instead, the system has to rely on measurements taken by system sensors, typically real valued, to perform the classification. In the rough set approach, the measurements are often converted into a finite set of discrete features called *attribute values*, which are then used to classify objects. The rough set theory does not make any assumptions how the classification is performed. It just assumes that some kind of classification knowledge and an associated classification procedure exist. In the rough set theory, the classification is represented in mathematical form by an equivalence relation, also referred to as an *indiscernibility relation*, denoted as IND , on the universe U , $IND \subseteq U \times U$. The relation is assumed to have a finite number of equivalence

classes, i.e., classification categories, usually called *elementary sets*. The collection of elementary sets of the IND relation is denoted as IND^* . If $IND^* = \{U\}$, that is, if the relation IND has only one elementary set equal to the universe U , then the classification is called *trivial*.

The combined structure consisting of the universe and its classification, reflecting our ability to distinguish individual objects, is called an *approximation space*. It is formally denoted as a pair (U, IND) in rough set theory.

Set Approximation Regions

The elementary sets reflect our basic knowledge about the domain, the knowledge in the sense of knowing both, which categories of objects exist in the application domain, and being able to classify each object into one of the categories. With this kind of knowledge, it is not possible, in general, to construct discriminating description of any arbitrary subset of the domain. What this limitation means in practice is that some concepts, corresponding to some subsets X of the universe U , can never be defined, or learned, with certainty with the available information. Such subsets are called *undefinable*, or *rough*. At best, only approximate, or uncertain, definitions of rough sets can be created, based on knowing, or learning from data, their set-theoretic relationship with respect to set inclusion and set overlap with elementary sets. Based on this relationship, Pawlak's rough set theory recognizes the following kinds of *approximation regions* and related set approximations (Pawlak 1982, 1991).

The *lower approximation* in the approximation space (U, IND) , or alternatively the *positive approximation region*, $POS(X)$, of an arbitrary subset, referred here as target set $X \subseteq U$, is defined as the union of those equivalence classes of IND^* which are completely contained in X :

$$POS(X) = \cup\{E \in IND^* : E \subseteq X\}. \quad (1)$$

The lower approximation characterizes objects which can be classified into X with *certainty*, based on the available target set versus elementary set relationship information. It is the largest set of objects which has discriminating description and is contained in the target set X .

If $POS(X) = X$, then the set X is said to be *definable*, or *precise*. A definable set can be expressed as a set union of some elementary sets forming the approximation space. In such a case, the discriminating description of the set X can be formed in terms of available information, and for each object of the universe U , it can be determined with certainty as to whether it belongs to X or not.

If $POS(X) \neq X$, then the set X is *undefinable*, or *rough*, in which case there are some objects in the universe whose membership status with respect to X cannot be determined based on the available information.

The *upper approximation* defines objects which can *possibly* belong to the target set. It is the smallest set of objects having discriminating description and containing the target set. In other words, the upper approximation is the smallest definable superset of the target set. The upper approximation of X , denoted here as $UPP(X)$, is a union of these elementary classes which have some set overlap with the set X :

$$UPP(X) = \cup\{E \in IND^* : E \cap X \neq \emptyset\}. \quad (2)$$

The quality of the approximation space used to form partial or full description of the target set of objects X is expressed in these definitions in terms of the ability to form tight lower and upper approximations of the target set. For a definable set, we have $X = UPP(X) = POS(X)$. In all other cases, we have proper inclusion $POS(X) \subset UPP(X)$, which leads to the following notion of the boundary region $BND(X)$:

$$BND(X) = UPP(X) - POS(X). \quad (3)$$

The boundary region is an area of total uncertainty. It is the largest definable set of objects all of which *possibly*, but not with certainty, belong to the target set X . The smaller the boundary area, the better the upper and lower approximations approximate the actual definition of the target set X . In practice, the actual definition of the target set is normally unknown. It is a subject of approximate learning from data, in the rough set context, in machine learning or data mining applications (Ziarko 2002b).

Outside of the upper approximation lies the *negative region*. The negative region is composed of the objects belonging to elementary sets which are known to be disjoint with the target set X . It is therefore known with certainty that they do not belong to the target set X . Formally, the negative region $NEG(X)$ is defined as a union of those elementary sets of IND^* which are completely disjoint with X :

$$NEG(X) = \cup \{E \in IND^* : E \cap X = \emptyset\}. \quad (4)$$

The negative region characterizes objects which can be classified into target set complement $\neg X$ with certainty, based on the available information. It is the largest set of objects which has discriminating description and is contained in the complement of the target set X .

Variable Precision Rough Set Model

There are many extensions and generalizations of the rough set approach, which inspired significant research interest. The primary motivation behind research aimed at extending rough set approach is to increase the scope of applications of rough set theory to deal with the imperfections of gathered application data. In particular, application data often suffer from the presence of measurement noise, leading to lack of consistency and resulting difficulty to form data classifications and set approximations, as required by Pawlak's rough set theory. One of the first extensions of the rough set theory along these lines was the variable precision rough set model (VPRSM), introduced by the author. Its original definition first appeared in (Ziarko 1993). It was subsequently extended to cover a broader scope of potential applications in the form of variable precision rough set model with asymmetric bounds (Katzberg and Ziarko 1994). Further refinements and additions were later incorporated. In this entry, the most recent formulation of VPRSM is presented.

Set Frequency and Conditional Frequency

In practical applications of rough set methodology, a common issue is the existence of an extensive

boundary area, which is not very useful in the context of creation of definitive prediction, or decision rules from data. This is often due to limited "resolution" of the adopted object classification procedure. It is not possible to draw any definitive conclusions about the relationship between boundary area objects and the target set since boundary area objects can only *possibly* belong to the target set, or to its complement. The frequency distribution information, implicitly present in the boundary area, is not used in the original formulation of rough sets.

The VPRSM methodology allows to classify arbitrary objects into the "goal" class, or its complement, with an error rate which is considered acceptable in the context of predefined criteria. The criteria are assumed to be domain specific and, consequently, outside of the scope of the rough set model. This approach forms a basis for development of approximate prediction, or decision system from data with a controlled degree of associated uncertainty. For example, the objective may be to predict (diagnose) the presence, or absence, of a specific disease based on the results of medical tests, which would increase the accuracy of such predictions (diagnoses) in comparison to predictions based solely on the frequency of occurrence of the disease in the population.

In the VPRSM approach, each equivalence class E of the indiscernibility relation IND is assigned two measures: the relative "size" of the class E within universe U , that is, the frequency $P(E)$, assumed $P(E) > 0$, and the relative "size" of the target set X within an elementary set E , that is, the conditional frequency $P(X|E)$. The conditional frequency, in this context, is just a measure of the degree of overlap between the target set X and the elementary set E . These two measures can be obtained from data respectively by:

$$P(E) = \frac{card(E)}{card(U)} \quad (5)$$

and

$$P(X|E) = \frac{card(X \cap E)}{card(E)}, \quad (6)$$

where *card* denotes set cardinality. If additional assumptions concerning randomness of data

selection are satisfied, the frequency measures can be treated as estimates of probabilities occurring in a domain from which data were selected. Although we do not make such assumptions, we will call the calculated frequencies as probabilities, or conditional probabilities, respectively.

The target set X may be undefinable (Pawlak 1982, 1991), or *rough*, which means that it cannot be expressed as a set union of some elementary sets forming the approximation space. That is, in general, the set-definability criterion:

$$X = \cup\{E \in IND^* : E \subseteq X\}. \quad (7)$$

is not satisfied.

This lack of definability is more common than not in applications. The VPRSM approach extends the rough set model to deal more effectively with undefinable sets. The VPRSM is replacing the full inclusion relationship between sets with the overlap relationship in the definitions of set approximations. Two precision control parameters, called lower limit l and upper limit u , are introduced in the definitions of upper, or lower set approximations, respectively. In this way, it is possible to control the process of creation of approximations of a target set to identify such approximations which satisfy user-imposed criteria, such as cost criteria, or characterizing classes of patients with an elevated (or reduced) risk of a disease.

Approximation Regions in the VPRSM

The notion of prior probability of the target set X , or frequency $P(X)$, with the assumption $0 < P(X) < 1$, plays an essential role in the definitions of set X approximations. If the available data were selected at random, it represents the likelihood that a random object $e \in U$ is a member of the target set X in the absence of any classification knowledge about the object. If the classification knowledge is available, as represented by any non-trivial indiscernibility relation IND , the likelihood of the membership in the target set X can either increase, or decrease, or stay approximately the same as the prior probability $P(X)$. These variations in the set X membership likelihood across different elementary sets are reflected in the definitions of set

approximation regions. They characterize areas of the universe U with *significantly* increased, *significantly* decreased, or approximately unchanged target set X membership probability.

The approximation regions in the VPRSM are defined in terms of two *precision control* parameters: the upper limit u and the lower limit l . Each elementary set can be classified into one of three approximation regions of the set X : a positive region $POS_u(X)$, a negative region $NEG_l(X)$, or a boundary region $BND_{l,u}(X)$.

The upper limit u defines the positive approximation region, or lower approximation, of the target set X , with the constraint $0 < P(X) < u \leq 1$. It represents the least acceptable degree of the conditional probability $P(X|E)$, or the set overlap degree, to include the elementary set E in the positive region. Here, the positive region $POS_u(X)$ of the target set X is a collection of objects for which the probability of membership in the target set X is *significantly* higher than the prior probability $P(X)$, where the term *significantly higher* is precisely specified by the parameter u :

$$POS_u(X) = \cup\{E : 0 < P(X) < u \leq P(X|E) \leq 1\}. \quad (8)$$

The lower limit l defines the negative approximation region of the target set X , with the associated constraint $0 \leq l < P(X) < 1$. It is the highest acceptable degree of the conditional probability $P(X|E)$ to include the elementary set E in the negative region. The negative region $NEG_l(X)$ of the target set X is a collection of objects for which the probability of membership in the target set X is *significantly* lower than the prior probability $P(X)$, where the term *significantly lower* is precisely specified by the parameter l :

$$NEG_l(X) = \cup\{E : 0 \leq P(X|E) \leq l < P(X) < 1\}. \quad (9)$$

The boundary approximation region, denoted as $BND_{l,u}(X)$, is a collection of remaining objects, the ones which cannot be classified with sufficient certainty into either positive or negative regions.

For the boundary area objects, the probability of membership in the target set X is not significantly different from its prior probability $P(X)$:

$$BND_{l,u}(X) = \cup\{E : l < P(X|E) < u\}. \quad (10)$$

In the Pawlak's rough set model (Pawlak 1982, 1991), the notion of upper approximation of a target set is defined as a union of all elementary sets which have nonempty intersection with the set. It is a union of positive and boundary regions. The generalized definition of upper approximation $UPP_l(X)$ in the VPRSM approach is also a set union of the positive region and of the boundary region, which can be expressed compactly as:

$$UPP_l(X) = \cup\{E : l < P(X|E) \leq 1\}. \quad (11)$$

The settings of the parameters l and u are entirely dependent on the requirements of a practical application, by being likely subjective or obtained via the cost-benefit analysis (Ziarko 1999), subject to the general constraint $0 \leq l < P(X) < u \leq 1$.

Note that, when $l = 0$, this generalized definition coincides with the Pawlak's definition of upper approximation. In addition, when $u = 1$, it can be easily demonstrated that the VPRSM definitions of positive, negative, and boundary regions become equivalent to the original rough set model's definitions of lower approximation, negative and boundary regions (Pawlak 1982, 1991). One can also note that, as opposed to Pawlak's rough sets, it is not, in general, true that $POS_u(X) \subseteq X$ and that $X \subseteq UPP_l(X)$.

Absolute Set Approximation Regions

Regardless of the choice of lower and upper limit precision control parameters, the positive and negative approximation regions are subsets of *absolute approximation regions*. They represent areas of the universe U characterized by an unconstrained increase, or decrease, of the set X membership probability. In this case, no parameters to specify "sufficiently" high increase, or

decrease, of the set membership probability in those areas are required. These areas of the universe are called *absolute approximation regions*.

The *absolute boundary region* $BND^*(X)$ of the target set X is a definable region of the universe U consisting of those elementary sets which are characterized by the unchanged probability of membership in the target set $X \subseteq U$ in relation to the prior probability:

$$BND^*(X) = \cup\{E : 0 < P(X|E) = P(X) < 1\}. \quad (12)$$

In the absolute boundary region, each elementary set E is probabilistically independent from the target set X , that is, $P(X \cap E) = P(X)P(E)$. Consequently, the whole boundary region is probabilistically independent from the target set X . In other words, objects in the absolute boundary region are entirely unrelated to the target set X .

The region of the universe U that is characterized by an increased probabilistic association with the target set $X \subseteq U$, relative to the prior probability $P(X)$, is called *the absolute positive region* of the set X , denoted as $POS^*(X)$:

$$POS^*(X) = \cup\{E : 0 < P(X) < P(X|E) \leq 1\}. \quad (13)$$

In the absolute positive region of X , the likelihood of an object belonging to the target set is higher than in the whole universe U .

Similarly, the *absolute negative region*, $NEG^*(X)$, of the target set X is an area of the universe U characterized by diminished likelihood of an object being a member of the target set X , in relation to its prior probability $P(X)$:

$$NEG^*(X) = \cup\{E : 0 \leq P(X|E) < P(X) < 1\}. \quad (14)$$

As in the previous section, the *absolute upper approximation* $UPP^*(X)$ of set X can be constructed as a union of positive and boundary regions, giving:

$$UPP^*(X) = \cup \{E : 0 < P(X) \leq P(X|E) \leq 1\}. \quad (15)$$

The above definitions of absolute approximation regions provide the basis for the Bayesian rough set model (Slezak and Ziarko 2003; Ziarko 2001).

Decision Tables Learned from Data

Pawlak introduced the notion of decision tables obtained from data in the framework of his rough set theory (Pawlak 1982, 1991). This notion was further extended in the context of VPRSM, to be called *probabilistic decision tables* (Ziarko 1999, 2002a, 2003; Chen and Ziarko 2010). Probabilistic decision tables describe elementary sets of approximation space and their probabilistic relations with a target set. They are composed of combinations of attribute values, probability values, and approximation region designations. They are useful for analysis of data dependencies, and also for machine learning and data mining applications (Ziarko 1999, 2003; Chen and Ziarko 2010). In machine learning applications, they can be learned from data by learning the classes of the approximation space and by estimating their probabilistic relationships with the target set.

Attributes

In many applications, the information about objects is expressed in terms of values of observations or sensor measurements, often real valued, referred to as *features*. For the purpose of rough set-based analysis and classifier construction, the feature values are typically mapped into finite-valued numeric or symbolic domains to form composite mappings, referred to as *attributes*. A common kind of mapping is dividing the range of values of a feature into a number of suitably chosen disjoint subranges via a discretization procedure (see Nguyen 2003). Formally, an attribute a is a function on the universe U , $a: U \rightarrow a(U) \subseteq V_a$, where V_a is a finite set of values called the *domain* of the attribute a .

Based on combinations of attributes and their values, a structure of approximation space can be

created and analyzed using general notions and results of rough set theory or of the VPRSM. Each attribute defines a classification of the universe U into classes corresponding to different values of the attribute. In more detail, each attribute value, $v \in a(U)$, corresponds to a set of objects $E_v^a \subseteq U$ such that $E_v^a = a^{-1}(v) = \{e \in U : a(e) = v\}$. The classes E_v^a , referred to as *a-elementary sets*, form a partition of the universe U . A partition of the universe U , and the corresponding approximation space, can also be created on the basis of any subset of attributes B , based on combinations of attribute values of attributes belonging to the subset B .

Deterministic and Nondeterministic Decision Tables

A *knowledge representation system* (Pawlak 1982, 1991) is a pair (U, A) , where U is a universe and A is a nonempty and finite set of attributes defined on U . In the context of rough set approach, decision tables are constructed in terms of knowledge representation systems as follows.

Let $C, D \subset A$ be two disjoint subsets of attributes, called *condition* and *decision* attributes, respectively. The condition attributes generate the partitioning of the universe U into classes of objects having identical values of attributes belonging to C , thus forming the structure of approximation space on U . The corresponding collection of elementary sets of this approximation space is denoted by U/C . Similarly, the decision attributes D induce a structure of approximation space on U , with U/D denoting its elementary sets. A tabular presentation of a knowledge representation system with defined condition and decision attributes is called a decision table (Pawlak 1991). Decision tables fall into two broad groups: *deterministic decision tables* and *nondeterministic decision tables*.

Deterministic decision tables describe the functional relation between a set of observations (inputs, conditions) and the corresponding decisions (outcomes). In practice, the deterministic decision knowledge, to create deterministic decision table, is not always available. When only some, but not all, decisions can uniquely be determined by combinations of condition attribute values, the decision table is called nondeterministic. In a nondeterministic

decision table, the relationship between conditions and decisions is either partially functional, or totally nonfunctional.

Probabilistic Decision Tables

Compared to the previous two types of decision tables, which are based on Pawlak's rough set theory, a probabilistic decision table is developed within the framework of the variable precision rough set theory. It contains some built-in probabilistic measures to help in the process of decision-making or prediction in nondeterministic decision situations.

In probabilistic decision tables, the focus is on elementary sets (called target sets here) of the decision attributes D , $X \in U/D$, of the partition generated by the decision attributes.

For a given target set X , the probabilistic decision table is a mapping associating each combination of condition attribute values, corresponding to an elementary set $E \in U/C$, with a triple of values representing:

1. The unique designation of the rough approximation region (positive, negative, or boundary region)
2. The respective values of the elementary set probability $P(E)$
3. The conditional probability $P(X|E)$

In practice, when deriving a probabilistic decision table, the measures of $P(E)$ and $P(X|E)$ are usually computed based on available data. An example of probabilistic decision table is shown in Table 1 with lower limit $l = 0.3$, upper limit $u = 0.8$, and the prior probability $P(X) = 0.4366$. Probabilistic decision tables can also be constructed in a similar way within the framework of Bayesian rough set model, based on absolute approximation regions.

Dependencies in Probabilistic Decision Tables

Dependencies between attributes of probabilistic decision tables play an important role in data analysis applications. In particular, the interest is

Variable Precision Approximations of Rough Sets, Table 1 Probabilistic decision table with $l = 0.3$, $u = 0.8$, and $P(X) = 0.4366$

a_1	a_2	a_3	Region	$P(E_i)$	$P(X E_i)$
1	1	1	BND	0.0520	0.78
1	1	0	NEG	0.1354	0.02
1	0	1	POS	0.1562	0.99
1	0	0	BND	0.1562	0.36
0	1	1	NEG	0.1406	0.11
0	1	0	BND	0.1093	0.41
0	0	1	NEG	0.1562	0.27
0	0	0	POS	0.0941	0.85

in dependencies occurring between condition attributes C , or their subset, and two-class classification $(X, \neg X)$ corresponding to the target set X and its complement $\neg X$. This classification is numerically represented in both, classification and probabilistic decision tables, by values of the conditional probability $P(X|E)$. Technically, the columns $P(E)$ and $P(X|E)$ can be treated as extra "attributes" associating some real values with elementary sets of the classification generated by condition attributes. In particular, the attribute $P(X|E)$ describes the distribution of the degrees of association across different elementary sets E with the target set X . Consequently, it can be used, in conjunction with the attribute $P(E)$, for computing the overall degree of association of the set of condition attributes, or of its subset, with the binary classification of the universe U , as defined by the target set X and its complement $\neg X$.

Two kinds dependencies, called γ - dependency and λ - dependency, provide useful measures for evaluating inter-attribute connections in probabilistic decision tables. They also provide criteria for decision table optimization through the reduction of redundant condition attributes.

γ - Dependency Measure

Functional dependencies and partially functional dependencies between attributes of decision tables were originally explored by Pawlak in (1982, 1991). We will refer to them as γ - dependencies. They capture the quality of approximation of the target set $X \in U/D$ in terms of the elementary sets of the approximation space

induced by condition attributes. We generalize them within the framework of the VPRSM by defining the γ – *dependency* (Ziarko 2006) as a relative size of the positive region of the two class partition $(X, \neg X)$, subject to prior setting of the values of the control parameters l and u :

$$\gamma_{l,u}(X|C) = P(POS_u(X|C) \cup NEG_l(X|C)), \quad (16)$$

where $POS_u(X|C)$ and $NEG_l(X|C)$, respectively, are positive and negative regions of X in the approximation space induced on U by the set of condition attributes C . This dependency measure reflects the proportion of objects in the universe U that can be classified as members of the target set X , or a complement of the target set X , with sufficient certainty, as given by the parameters l and u . The measure $\gamma_{l,u}(X|C)$ can be computed directly from the probabilistic decision table by using:

$$\gamma_{l,u}(X|C) = \sum_{E \subseteq POS_u(X|C) \cup NEG_l(X|C)} P(E). \quad (17)$$

The $\gamma_{l,u}(X|C)$ measure was inspired by the partial functional dependency measure introduced by Pawlak (1982, 1991). The partial functional dependency was defined as a fraction of objects of the universe U that can be uniquely classified, based on their condition attribute value combinations, as members of some classes of the decision attribute D . More precisely, in the VPRSM terms, the dependency $\gamma(D|C)$ is given by:

$$\gamma(D|C) = \sum_{F \in U/D} P(POS_1(F|C)). \quad (18)$$

An important property of γ – *dependencies* is their *monotonicity* with respect to the addition of extra condition attributes. It is summarized in the following Monotonicity Theorem 1 for γ – *dependencies*:

Monotonicity Theorem 1 *Let $B \subseteq C$ be a subset of condition attributes on U , and let “ a ” be any condition attribute. Then the following relation holds:*

$$\gamma(X|B) \leq \gamma(X|B \cup \{a\}). \quad (19)$$

The Theorem 1 in essence means that expanding the collection of condition attributes does not result in a decrease of the γ – *dependency*, or vice versa, reducing the condition attributes will not result in an increase of the γ – *dependency*. The above Theorem 1 plays a useful role in the analysis and optimization of deterministic and nondeterministic decision tables, using the idea of γ – *reduct*, as described in section “Reduction and Significance of Attributes.”

λ – *Dependency Measure*

Another kind of dependency, which is unrelated to the γ – *dependency* measure, is a parametric λ – *dependency*, denoted as $\lambda_{l,u}(X|C)$ (Ziarko 2006). It represents the average, or expected, degree of the probabilistic connection between elementary sets E ($E \in U/C$) and the binary classification $(X, \neg X)$ corresponding to the target set X and its complement $\neg X$. The dependency is defined as a normalized expected degree of deviation of the conditional probability $P(X|E)$ from the prior probability $P(X)$:

$$\lambda_{l,u}(X|C) = \frac{\sum_{E \subseteq POS_u(X|C) \cup NEG_l(X|C)} P(E)|P(X|E) - P(X)|}{2P(X)(1 - P(X))}, \quad (20)$$

where $2P(X)(1 - P(X))$ is a normalization factor equal to the theoretically maximum value of the numerator summation, achievable only when X is definable in Pawlak’s rough set’s sense, independent of settings of the parameters l and u . The higher the deviation, the stronger the probabilistic connection between conditional attributes C and the decision partition $(X, \neg X)$, and vice versa, with the total probabilistic independence occurring at $\lambda_{l,u}(X|C) = 0$.

The $\lambda_{l,u}(X|C)$ dependency measure can be computed directly from the probabilistic decision table by computing the prior probability using:

$$P(X) = \sum_{E \in U/C} P(E)P(X|E). \quad (21)$$

In the framework of the Bayesian rough set model, the parametric λ – dependency reduces to nonparametric λ – dependency defined as:

$$\lambda(X|C) = \frac{\sum_{E \in U/C} P(E) | P(X|E) - P(X) |}{2P(X)(1 - P(X))}. \quad (22)$$

The parametric dependency λ – dependency measure and the nonparametric λ – dependency represent a normalized expected degree of deviation of the conditional probability $P(X|E)$ from the prior probability $P(X)$. The main practical advantage of the nonparametric λ – dependency is the absence of any external parameters, which may be difficult to obtain, to compute the dependency. As γ – dependencies, λ – dependencies are also monotonic with respect to condition attributes, as summarized in the following Monotonicity Theorem 2 for λ – dependencies (Slezak and Ziarko 2003):

Monotonicity Theorem 2 *Let $B \subseteq C$ be a subset of condition attributes on U , and let “ a ” be any condition attribute. Then the following relation holds:*

$$\lambda(X|B) \leq \lambda(X|B \cup \{a\}). \quad (23)$$

The Monotonicity Theorem 2 for λ – dependencies is particularly useful in the analysis and optimization of probabilistic decision tables by applying the notion of λ – reduct, as presented in the next section.

It should be noted that the above dependency measures, as presented for two class decision attribute, can naturally be extended to probabilistic decision tables with multiclass decision attributes.

Reduction and Significance of Attributes

In data analysis, data mining, or machine learning applications, one of the fundamental issues is finding a minimum subset, or subsets, of condition attributes which preserve the dependency between all condition attributes and the decision attribute. It is also important to evaluate the contribution of individual attributes to the dependency in question.

Attribute Reduct

The minimum subset of condition attributes maintaining a dependency is called *attribute reduct*. For example, in machine learning control applications, finding different reducts would be equivalent to identifying alternative groups of sensors to achieve a particular control task. In a medical application, it would mean finding alternative groups of medical tests and symptoms to produce a particular diagnosis. The general idea of the reduct was introduced by Pawlak in the context of γ – dependencies between attributes (Pawlak 1982, 1991). The concept of the reduct generated considerable amount of research interest, primarily as a method for feature selection (Slezak and Ziarko 2003; Zhang et al. 2003a, b; Zhong and Dong 2001; Mi et al. 2010; Inuiguchi et al. 2009; Xia Wang and Zhang 2008; Yao et al. 2006; Zhao et al. 2007; Yao and Zhao 2009; Beynon 2001; Swiniarski and Skowron 2003; James 2007; Bac and Tuan 2005). The notion of reduct in the framework of VPRSM is applicable to analysis and optimization of decision tables and probabilistic decision tables.

Formally, the γ – reduct of attributes with respect to γ – dependency is a minimal subset $RED \subseteq C$ of condition attributes preserving the γ – dependency with the target classification $(X, \neg X)$. That is, by definition, the γ – reduct, RED , of attributes satisfies the following two properties:

$$\gamma(X|RED) = \gamma(X|C) \quad (24)$$

and for any attribute $a \in RED$:

$$\gamma(X|RED - \{a\}) < \gamma(X|RED). \quad (25)$$

The Monotonicity Theorem 1 for γ – dependencies ensures the existence of at least one reduct with respect to γ – dependency and provides for a simple linear-time procedure for computation of such a reduct.

As a consequence of the Monotonicity Theorem 2 for λ – dependencies, the notion of the *probabilistic reduct*, or λ – reduct, of attributes $RED \subseteq C$ can also be defined (Slezak and Ziarko 2003). It is a minimal subset of attributes preserving the λ – dependency with the target classification $(X, \neg X)$ (Slezak and Ziarko 2003):

$$\lambda(X|RED) = \lambda(X|C) \quad (26)$$

and for any attribute $a \in RED$:

$$\lambda(X|RED - \{a\}) < \lambda(X|RED). \quad (27)$$

The probabilistic reducts can be computed using any method available for γ – reduct computation. As with dependency measures, the above notion of attribute reduct can be extended to probabilistic decision tables with multiclass decision attributes.

Significance of Attributes

Once a reduct, either a γ – reduct or λ – reduct, is computed, it is possible to analyze the contribution of individual reduct attributes to the dependency in question. They can be evaluated with respect to their contribution to the dependency with the target classification by adopting the notion of attribute *significance factor*. The significance factor with respect to λ – dependency, $sig_{RED}(a)$ of an attribute $a \in RED$ is a relative decrease of the dependency $\lambda(X|RED)$ caused by removal of an attribute “ a ” from the reduct:

$$sig_{RED}(a) = \frac{\lambda(X|RED) - \lambda(X|RED - \{a\})}{\lambda(X|RED)} > 0. \quad (28)$$

Similarly, the significance of attributes within any γ – reduct of a decision table can be evaluated by employing the same idea. The significance factor with respect to γ – dependency, $sig_{RED}(a)$ of an attribute $a \in RED$ is a relative decrease of the dependency $\gamma(X|RED)$ caused by removal of the attribute “ a ” from the reduct:

$$sig_{RED}(a) = \frac{\gamma(X|RED) - \gamma(X|RED - \{a\})}{\gamma(X|RED)} > 0. \quad (29)$$

In addition to evaluating the significance of reduct attributes, it is possible to identify the set of most essential condition attributes with respect to the dependency between attributes. These

attributes, which in the case of λ – dependency are called the λ – core attributes, would never be eliminated in the process of any λ – reduct computation. They are included in all λ – reducts, which means that their collection is equal to the intersection of all λ – reducts.

Any core attribute $\{a\}$ satisfies the following inequality:

$$\lambda(X|C) > \lambda(X|C - \{a\}). \quad (30)$$

The above inequality demonstrates that there is no need to compute all λ – reducts, which are NP-hard, to identify the λ – core. All core attributes can be found by simple linear testing procedure based on the above inequality.

As in the case of λ – core attributes, γ – core (Pawlak 1982, 1991) attributes can also be computed in a decision table with respect to the dependency $\gamma(X|C)$ by testing the effect of removal of each condition attribute.

Final Remarks

This entry reviews some key results of our research on fundamentals of the variable precision rough set model with its application to data-based acquisition, discovery and analysis of data dependencies, probabilistic decision table reduction, and analysis. The basics of the original Pawlak’s rough set theory and of the Bayesian rough set model are also discussed. The variable precision rough set approach was used in many applications since its original introduction in the 1990s. Potential application areas for presented methodology are ones dealing with large amounts of data such as sensor-based automated control, medicine, pattern classification, market analysis and prediction, machine learning, and data mining in general.

Bibliography

- Bac LH, Tuan NA (2005) Using rough set in feature selection and reduction in face recognition problem. In: PAKDD 2005, LNAI 3518, pp 226–233
- Beynon MJ (2001) Reducts within the variable precision rough sets model: a further investigation. Eur J Oper Res 134(3):592–605

- Beynon M, Peel MJ (2001) Variable precision rough set theory and data discretization: an application to corporate failure prediction. *Int J Manag Sci* 29:561–576
- Chen X, Ziarko W (2010) Rough set-based incremental learning approach to face recognition. In: *RSTC 2010*, LNAI 6086, pp 356–365
- Greco S, Matarazzo B, Slowinski R (2002) Multicriteria classification by dominance-based rough set approach. In: Kloesgen W, Zytlow J (eds) *Handbook of data mining and knowledge discovery*. Oxford University Press, New York
- Inuiguchi M, Yoshioka Y, Kusunoki Y (2009) Variable-precision dominance-based rough set approach and attribute reduction. *Int J Approx Reason* 50: 1199–1214
- James FP, Sheela R (2007) Feature selection: near set approach. In: *MCD 2007*, pp 57–71
- Katzberg JD, Ziarko W (1994) Variable precision rough sets with asymmetric bounds. In: Ziarko W (ed) *Rough sets, fuzzy sets and knowledge discovery*. Springer, London, pp 167–177
- Mi JS, Leung Y, Wu WZ (2010) Approaches to attribute reduction in concepts lattices. *Knowl Based Syst J* 23(6):504–511
- Nguyen HS (2003) On exploring soft discretization of continuous attributes. In: Pal SK, Polkowski L, Skowron A (eds) *Rough-neural computing: techniques for computing with words*. Cognitive Technologies. Springer, Berlin, pp 333–350
- Pawlak Z (1982) Rough sets. *Int J Comput Inf Sci* 11: 341–356
- Pawlak Z (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Dordrecht
- Slezak D, Ziarko W (2003) Attribute reduction in the Bayesian version of variable precision rough set model. *Electron Notes Theor Comput Sci* 82:263–273
- Swiniarski RW, Skowron A (2003) Rough set methods in feature selection and recognition. *Pattern Recogn Lett* 24(6):833–849
- Wei LL, Zhang WX (2004) Probabilistic rough sets characterized by fuzzy sets. *Int J Uncertain Fuzziness Knowl Based Syst* 12:47–60
- Xia Wang X, Zhang W (2008) Relations of attribute reduction between object and property oriented concept lattices. *J Knowl Based Syst* 21(5):398–403
- Yao YY (2007) Decision theoretic rough set models. In: *Proceedings of the Rough sets and knowledge, RSKT 2007*, LNAI 4481, pp 1–12
- Yao YY (2008) Probabilistic rough set approximations. *Int J Approx Reason* 49(2):255–271
- Yao YY, Lin TY (1996) Generalization of rough sets using modal logic. *Intell Autom Soft Comput* 2(2):103–120
- Yao YY, Yao BX (2012) Covering-based rough set approximations. *Inf Sci* 200:91–107
- Yao YY, Zhao Y (2009) Discernibility matrix simplification for constructing attribute reducts. *Inf Sci* 179(5): 867–882
- Yao YY, Zhao Y, Wang J (2006) On reduct construction algorithms. In: *Proceedings of the first international conference Rough sets and knowledge technology, RSKT 2006*, LNAI, vol 4062, pp 297–304
- Zhang J, Wang J, Li D, He H, Sun J (2003a) A new heuristic reduct algorithm base on rough sets theory. *Lect Notes Comput Sci* 2762:247–253
- Zhang WX, Mi JS, Wu WZ (2003b) Approaches to knowledge reductions in inconsistent systems. *Int J Intell Syst* 18(9):989–1000
- Zhang HY, Leung Y, Zhou L (2013) Variable precision-dominance based rough set approach to interval-valued information systems. *Inf Sci* 244:75–272
- Zhao Y, Luo F, Wong SKM, Yao YY (2007) A general definition of an attribute reduct. In: *Proceedings of the Second international conference Rough sets and knowledge technology, RSKT 2007*, LNAI, vol 4481, pp 101–108
- Zhong N, Dong J (2001) Using rough sets with heuristics for feature selection. *J Intell Inf Syst* 16:199–214
- Ziarko W (1993) Variable precision rough sets model. *J Comput Syst Sci* 46(1):39–59
- Ziarko W (1999) Decision making with probabilistic decision tables. In: *Proceedings of the 7th international workshop on rough sets, fuzzy sets, data mining and granular computing, RSFDGrC99*, Yamaguchi, Japan, LNAI 1711. Springer, Berlin, pp 463–471
- Ziarko W (2001) Set approximation quality measures in the variable precision rough set model. In: *Soft computing systems, management and applications*. IOS Press, Amsterdam, pp 442–452
- Ziarko W (2002a) Probabilistic decision tables in the variable precision rough set model. *Comput Intell* 17(3): 593–603
- Ziarko W (2002b) Rough set approaches for discovery of rules and attribute dependencies. In: Kloesgen W, Zytlow J (eds) *Handbook of data mining and knowledge discovery*. Oxford University Press, New York, pp 328–339
- Ziarko W (2003) Acquisition of hierarchy structured probabilistic decision tables, and rules from data. *Expert systems. Int J Knowl Eng Neural Netw* 20(5):305–310
- Ziarko W (2005) Probabilistic rough sets. *Lect Notes Comput Sci* 3641:283–293
- Ziarko W (2006) Partition dependencies in hierarchies of probabilistic decision tables. In: *RSKT 2006*, LNAI 4062, pp 42–49
- Ziarko W (2008a) Probabilistic dependencies in linear hierarchies of decision tables, transactions on rough sets 9. *Lect Notes Comput Sci* 5390:444–454
- Ziarko W (2008b) Probabilistic approach to rough sets. *Int J Approx Reason* 49(2):272–284

Algebraic Models and Granular Computing

Anita Wasilewska
Computer Science Department, Stony Brook
University, Stony Brook, NY, USA

Article Outline

Glossary
Definition of the Subject
Introduction: Algebraic Models
Granular Computing
Topological Rough Algebras
Generalized Rough Sets, L-Fuzzy, and
LT-Fuzzy Sets
Bibliography

Glossary

Syntax Syntax, or syntactical concepts refer to simple relations among symbols and formulas of formal, symbolic languages. The formal languages, even if created with a specific meaning in mind, do not carry themselves any meaning. The meaning is being assigned to them by establishing a proper semantics.

Semantics Semantics for a given symbolic language $\mathcal{L}$ assigns a specific interpretation in some domain to all symbols and expressions of the language. It also involves related ideas such as truth and model. They are called semantical concepts to distinguish them from the syntactical ones.

Algebraic Model The word *model* is used in many situations and has many meanings but they all reflect some parts, if not all, of its following intuitive definition. A structure M is a model for a set $\mathcal{F}$ of formulas of a formal language $\mathcal{L}$ if and only if every formula $A \in \mathcal{F}$ it true in M . When the notion of truthfulness is established in terms a certain abstract algebra M is called an *algebraic model*.

Abstract Algebra By *abstract algebra*, or shortly an *algebra* we mean any system. $\mathbf{A} = (A, o_1, o_2, \dots, o_n)$, where A is a nonempty set and o_i are k argument function defined on A with values in A . They are called algebra k argument operations, for certain $k \geq 0$.

Boolean Algebra An abstract algebra $(A, 1, 0, \Rightarrow, \cap, \cup, \neg)$ is a Boolean algebra if it is a distributive lattice with the smallest element 0 and the greatest element 1 and every element $a \in A$ has a complement $\neg a \in A$. Boolean algebras are algebraic models for classical logic.

Topological Boolean Algebra By a topological Boolean algebra we mean an abstract algebra $(A, 1, 0, \Rightarrow, \cap, \cup, \neg, I)$ where $(A, 1, 0, \Rightarrow, \cap, \cup, \neg)$ is a Boolean algebra and, moreover, the following conditions hold: $I(a \cap b) = Ia \cap Ib$, $Ia \cap a = Ia$, $I Ia = Ia$, and $I 1 = 1$, for any $a, b, \in A$. Topological Boolean Algebras are algebraic models for some modal logics.

Algebra Representation Theorem Algebras are fairly abstract structures. Their deeper mathematical meaning is established by a proper *representation theorem*. The representation theorem establishes a relationship between logic, algebra, set theory, and topology. The existence of the representation theorem is always the first question one asks about a newly created algebra.

Rough Set The starting point of the rough set theory is the assumption that we have initially some information (knowledge) about the elements of the universe U and that some elements can be indiscernible (equivalent) in the context of the information about them, that they can be “seen” as the same. This indiscernibility is defined by an equivalence relation R on U , and for any set $A \subseteq U$ we define, its lower and upper approximations $\underline{A}$ and $\overline{A}$. The set A is observed *precisely* only if $\underline{A} = \overline{A}$. Otherwise A is observed only “*roughly*,” and is called a *rough set*.

Rough Equality Given an equivalence relation R on a set U . Let $A \subseteq U$ and $\underline{A}, \overline{A}$ be the lower and upper approximations of the set A determined by R , respectively. We say that the sets A and B are *roughly equal* and denote it by $A \sim_R B$ if and only if $\underline{A} = \underline{B}$ and $\overline{A} = \overline{B}$.

Topological Quasi-Boolean Algebras An algebra $(A, \cap, \cup, \sim, 1, I)$ is called a *topological quasi-Boolean algebra* if $(A, \cup, \cap, \sim, 1)$ is a quasi-Boolean algebra and for any $a, b \in A$, $I(a \cap b) = Ia \cap Ib$, $Ia \cap a = Ia$, $Ia = Ia$, and $I1 = 1$.

Topological Rough Algebras A topological quasi-Boolean algebra $(A, \cap, \cup, \sim, I, C, 0, 1)$ such that for any $a \in A$, $C Ia = Ia$, is called a *topological rough algebra*. Topological rough algebras provide a deeper mathematical meaning to the rough sets theory in general, and to the notion of rough equality in particular.

Definition of the Subject

We use the algebraic models techniques to link rough set theory to logic, abstract algebras and topology.

In particular, we show that the notion of *rough equality* of sets leads, via algebraic models constructions, to a definition of new classes of algebras, called here *topological quasi-Boolean algebras* and *topological rough algebras*. These algebras generalize McKinsey and Tarski notion of closure algebras and the notion of quasi-Boolean algebras.

We formulate and prove the representation theorems for these two, rough sets theory inspired classes of algebras.

A concept of LT-fuzzy sets was introduced by Rasiowa and Cat Ho (1992). LT-fuzzy sets are a modification of L-fuzzy sets introduced by Goguen (1967). We introduce here a notion of a generalized rough set and show that it can be considered as a particular case of an L-fuzzy set. We also generalize the notion of rough equality of sets, introduced in Novotny and Pawlak (1985) to a notion of topological and rough topological equality of sets. We show that LT-fuzzy sets

provide a common characterization for rough, generalized rough sets and topological and rough equality congruent spaces.

Introduction: Algebraic Models

It is well known that sometimes a theory designated originally as a tool for the study of a practical problem generates a purely mathematical interest. In such case the theory becomes generalized way beyond the point needed for applications, the generalizations make use of other theories, often in completely unexpected ways, and the subject becomes established as a new part of pure mathematics. Physics is not the only such external source of mathematical theories. Another, somewhat surprising addition is discovery of *algebraic models* for classical and various nonclassical logics and the development of a branch of pure mathematics that it has brought to life, called algebraic logic.

The algebraic logic starts from purely logical considerations, abstracts from them, places them into a general contest of algebraic models, and makes use of other branches of mathematics such as topology, set theory, and functional analysis. It provides tools to discuss the modern symbolic logic from a different perspective. It serves to clarify its problems and to establish their connection with ordinary mathematics.

It is difficult to establish who was the first to use the algebraic models and methods. The investigations in logic of Boole himself led to the notion which we now call Boolean algebra, but one of the turning points in the algebraic study of logic was the introduction by Lindenbaum and in a slight different form by Tarski (1935) of a method of treating formulas, or equivalence classes of formulas as elements of an abstract algebra, called now the *Lindenbaum-Tarski algebra*.

A. Lindenbaum was a prominent Polish mathematician who was killed by the Nazis during the Second World War, and whose various results were not published. It was McKinsey, a close collaborator of Tarski, who first called the method of treating the set of all formulas of propositional logic “an unpublished method of Lindenbaum

explained to me by Professor Tarski” (McKinsey 1941).

The main use of the Lindenbaum-Tarski algebra is to show the correspondence between logic and abstract algebras. The algebra corresponding to classical logic is, of course, Boolean algebra.

Boolean and other algebras are fairly abstract structures. Their deeper mathematical meaning is established by a proper *representation theorem*. The existence of the representation theorem is always the first question one asks about a newly created algebra. The first representation theorem for Boolean algebras was established by Stone (1937) who stated the following:

Every Boolean algebra A is isomorphic to a field of sets. More exactly, A is isomorphic with a field of subsets of its Stone space S , which is a compact, totally disconnected Hausdorff space.

The representation theorem, hence, establishes a relationship between logic, algebra, set theory, and topology.

Other nonclassical logic algebraic models studies were initiated by Stone (1934), Tarski (1938), and McKinsey and Tarski (1948), followed by Henkin (1950) and Rasiowa and Sikorski (1953). The investigations of Tarski, McKinsey, Henkin, and Rasiowa led to what is now called algebraic models for intuitionistic, modal S4 and S5 logics, as opposed to Kripke models which were invented some 20 years later and applied to the intuitionistic and a variety of modal logics in Kripke (1965). The development of algebraic models lead to giving a deeper meaning of completeness of a given logic. For example, the completeness theorem for the classical propositional logic becomes exactly the same as Stone’s representation for Boolean algebras, the Gödel completeness theorem for the classical predicate logic is obtained as a result of the Baire theorem on sets of the first category in topological spaces. This was proved by Rasiowa and Sikorski (1953) as the part of the first algebraic proof of the Gödel completeness theorem ever given.

Other algebraic proofs were later given by Rieger (1951), Beth (1951), Łoś (1951), Reichbach (1955), and Scott (1954). The further algebraic considerations, undertaken by many

authors, produced a lot of important and interesting results like Łoś’s concept of an ultraproduct of models and its applications (Łoś 1954); the cylindric algebras of Henkin, Tarski, and Thompson (Henkin 1955a; Henkin and Tarski 1961; Tarski and Thompson 1952); Halmos’s theory of polyadic algebras (Halmos 1962); and finally, Tarski’s theory of prime filters in Boolean algebras (Tarski 1954) and a simple algebraic proof by Scott of the independence of continuum hypothesis.

A uniform, systematic presentation of the algebraic approach to classical, intuitionistic, and modal S4 and S5 propositional and predicate logics is provided by Rasiowa and Sikorski in their classic book *The Mathematics of Metamathematics* (Rasiowa and Sikorski 1963).

Further studies of nonclassical logics from the algebraic point of view followed. The book *An Algebraic Approach to Non-Classical Logics* (Rasiowa 1974) formulates an algebraic approach to a carefully selected widest possible class of logics which can be approached from algebraic point of view. The basic concept of the book is that of an implicative algebra. All logics belonging to the class mentioned above have algebraic models that are extensions of this algebra. The set of elements of any algebra considered in the book is ordered by a relation determined by the operation corresponding to implication and a zero-argument operation in that algebra, that is, by a constant element of it. This element is the greatest element with respect to the implication determined ordering. The algebras discussed in the book are the following: positive implication algebras, implication algebras, lattices and distributive lattices, quasi-Boolean algebras, relatively pseudo complemented, contrapositionally complemented and semi-complemented lattices, pseudo-Boolean algebras, quasi-pseudo-Boolean algebras, Boolean and topological Boolean algebras, and finally, Post algebras. The quasi-Boolean algebras, quasi-pseudo-Boolean algebras ($\mathcal{N}$ – lattices), and Post algebras have not been discussed in any previous books.

The above algebras form algebraic models for corresponding logics, discussed in the second part of the book. These logics are positive implicative logic, classical implicative logic, minimal logic,

positive logic with semi-negation, constructive logic with strong negation, and Post m -valued logics.

The algebraic methods for nonclassical logics follow closely, overcoming their own difficulties, the patterns established for the classical logic. We present here a short overview of the algebraic set up for the classical logic.

Example: Boolean Algebras and Classical Logic

The relationship between Boolean algebras and logic is not surprising because its notion was created as the result of Boole's investigations of the structure of logic and he called his theory the *algebra of logic*. We present now, as an example, the algebraic proof of Gödel's completeness theorem for the classical propositional logic: *A formula A is provable if and only if it is a tautology*.

We define the fundamental notion of the Lindenbaum-Tarski algebra (see Definition 5) and use it to show how the completeness theorem is translated into the language of the theory of Boolean algebras. Its equivalent Boolean formulation is that every nonzero element of the Lindenbaum-Tarski algebra belongs to a maximal filter. Hence we obtain the completeness theorem from the theorem of existence of maximal filters in Boolean algebras or from the fundamental representation theorem stating that every Boolean algebra is isomorphic to a field of sets.

The role played by the set of axioms and rules of inference of the propositional calculus in that proof is reduced to showing that the Lindenbaum-Tarski algebra is a Boolean algebra.

It was shown in Henkin (1955b) and Łoś (1957) that the representation theorem for Boolean algebras can also be deduced directly from the completeness theorem formulated in a little stronger form.

This proves that the completeness theorem for the classical propositional logic and the representation theorem for Boolean algebras express the same mathematical content formulated in different languages.

Algebraic Proof of the Completeness Theorem

Let's consider a propositional language $\mathcal{L}$ containing a standard set $\{\cup, \cap, \Rightarrow, \neg\}$ of logical connectives.

Let $S = \{\mathcal{L}, C\}$ be a deductive system with a consequence operation given by a certain set of logical axioms AX and some rules of inference. We write

$$\vdash A$$

to denote that a formula A has a formal proof in S .

Observe (Łukasiewicz 1929; Tarski 1935) that the set F of formulas of $\mathcal{L}$ determines an abstract algebra

$$\mathcal{F} = (F, \cup, \cap, \Rightarrow, \neg), \quad (1)$$

whereby performing an operation on a formula (two formulas) means writing the formula having this operation as a main connective. For example

$$\cap(A, B) = (A \cup B).$$

We define binary relations $\leq$ and $\approx$ in the algebra $\mathcal{F}$ of formulas of $\mathcal{L}$ as follows. For any $A, B \in F$,

$$A \leq B \text{ if and only if } \vdash (A \Rightarrow B), \quad (2)$$

$$A \approx B \text{ if and only if } \vdash (A \Rightarrow B) \text{ and } \vdash (B \Rightarrow A). \quad (3)$$

In order for the proof system S to be called a classical logic its set of axioms and rules of inferences must be such that we are able to prove all facts included below.

- F1** The relation $\leq$ defined by (2) is a quasi-ordering in F .
- F2** The relation $\approx$ defined by (3) is an equivalence relation in F .

The equivalence class containing a formula A is denoted by $[A]$.

The quasi ordering $\leq$ in F as defined by (2) induces an ordering relation in $F/\approx$ defined as follows:

$$[A] \leq [B] \text{ if and only if } A \leq B, \text{ i.e. if } \vdash (A \Rightarrow B). \quad (4)$$

F3 The relation $\approx$ defined by (3) is a congruence in the algebra $\mathcal{F}$ of formulas defined by (1).

Lindenbaum-Tarski Algebra

The algebra

$$\mathcal{LT} = (F/\approx, \cup, \cap, \Rightarrow, \neg), \quad (5)$$

where the operations $\cup, \cap, \Rightarrow$, and $\neg$ are determined by the congruence relation (3) that is,

$$[A] \cup [B] = [(A \cup B)], \quad [A] \cap [B] = [(A \cap B)], \\ [A] \Rightarrow [B] = [(A \Rightarrow B)], \quad \neg[A] = [\neg A].$$

is called a Lindenbaum-Tarski algebra of S .

F4 The Lindenbaum-Tarski algebra $\mathcal{LT} = (F/\approx, \cup, \cap, \Rightarrow, \neg)$ of S is a Boolean algebra. The unit element 1 is the greatest element in $(F/\approx, \leq)$, where the order relation $\leq$ is defined by (4). Moreover, for any formula A of S ,
 $\vdash A$ if and only if $[A] = 1$.

Algebraic Proof of the Completeness Theorem

In order to carry on the algebraic proof of the completeness theorem we additionally need the following notions and theorems concerning Boolean algebras.

Let $\mathbf{B} = (B, 1, \cup, \cap, \Rightarrow, \neg)$ be a Boolean algebra.

Filter A non-empty set Δ of elements of the algebra $\mathbf{B}$ is said to be a *filter* in $\mathbf{B}$ provided that, for any elements $a, b \in B$,
 $a \cap b \in \Delta$ if and only if $a \in \Delta$ and $b \in \Delta$;

Ideal A non-empty set Δ of elements of the algebra $\mathbf{B}$ is said to be an *ideal* in $\mathbf{B}$ provided that, for any elements $a, b \in B$,
 $a \cup b \in \Delta$ if and only if $a \in \Delta$ and $b \in \Delta$.

Maximal A filter (ideal) is called **maximal** in $\mathbf{B}$ if it is proper and is not any proper subset of a proper filter (ideal) in $\mathbf{B}$.

F5 Let $\mathbf{B} = (B, 1, \cup, \cap, \Rightarrow, \neg)$ be a Boolean algebra with the unit element 1 and zero element 0. For every element $a \neq 0$ ($a \neq 1$) in $\mathbf{B}$ there exists a maximal filter ∇ (maximal ideal Δ) in $\mathbf{B}$ such that $a \in \nabla$ ($a \in \Delta$).

Equivalence determined by ∇

Let ∇ be a filter in a Boolean algebra $\mathbf{B}$. We define, for any $a, b \in B$,

$$a \approx_{\nabla} b \text{ if and only if } a \Rightarrow b \in \nabla \text{ and } b \Rightarrow a \in \nabla.$$

The relation $\approx_{\nabla}$ is an equivalence relation, called the equivalence relation **determined** by the filter ∇ . We denote the equivalence classes of the equivalence relation $\approx_{\nabla}$ by

$$[a]_{\nabla}.$$

F6 Let $\mathbf{B} = (B, 1, \cup, \cap, \Rightarrow, \neg)$ be a Boolean algebra. The equivalence relation $\approx_{\nabla}$ determined by the filter ∇ is a congruence in $\mathbf{B}$ and the quotient algebra
 $B/\nabla = (B/\nabla, 1, \cup, \cap, \Rightarrow, \neg)$ is a Boolean algebra.

It follows immediately from de Morgan Laws that if ∇ is a filter in a Boolean algebra $\mathbf{B}$, then the set of all elements $\neg a$ where $a \in \nabla$ is an ideal called the *adjoint ideal* of ∇ . Assigning to every filter ∇ its adjoint ideal Δ , we establish a natural one-to-one correspondence between the class of all filters in $\mathbf{B}$ and all ideals in $\mathbf{B}$. In particular, if Δ is an ideal in $\mathbf{B}$, we can also form a quotient algebra

$$\mathbf{B}/\Delta.$$

By definition, $\mathbf{B}/\Delta$ is the quotient algebra $\mathbf{B}/\nabla$, where ∇ is the filter adjoint to Δ . We denote the elements of $\mathbf{B}/\Delta$ by

$$[a]_{\Delta}.$$

By this correspondence, it makes no practical difference from the practical point of view whether we talk about filters or ideals in Boolean algebras. This duality fails for many other algebras and hence from the practical applications to non-classical logics we introduce, as the next example, the notion of an implicative filter. In the case of Boolean algebras both notions coincide.

F7 The following conditions are equivalent for every filter ∇ (ideal Δ) of a Boolean algebra $\mathbf{B}$:

- (i) the filter ∇ (ideal Δ) is maximal;
- (ii) the quotient algebra $\mathbf{B}/\nabla$ ($\mathbf{B}/\Delta$) is a two-element Boolean algebra and the following conditions hold.

$$\begin{aligned} a \in \nabla \text{ if and only if } [a]_{\nabla} &= 1; \\ a \in \Delta \text{ if and only if } [a]_{\Delta} &= 0 \end{aligned}$$

Theorem 1 (Completeness Theorem)

If the **sound** propositional deductive system S is such that the Lindenbaum-Tarski algebra $\mathcal{LT}$ of S is a Boolean algebra, then for any formula A of S ,

$$\vdash A \text{ if and only if } \models A.$$

Proof As the system S is sound we have to show only that $\models A$ implies $\vdash A$. Suppose $\not\vdash A$. By **F4**, $[A] \neq 1$. By **F5**, there exists a maximal ideal Δ in the Lindenbaum-Tarski algebra $\mathcal{LT}$ of S such that $[A] \in \Delta$.

Let h be a natural homomorphism from $\mathcal{LT}$ onto the two element Boolean algebra $\mathcal{LT}/\Delta$. By **F7**, $h([A]) = [[A]]_{\Delta} = 0$.

Let VAR be the set of propositional variables of $\mathcal{L}$. We define

$$v : VAR \rightarrow F / \approx$$

by setting, for any $a \in VAR$, $v(a) = [a]$. Let v^* be the extension of v to the homomorphism of the algebra $\mathcal{F}$ of formulas and the Lindenbaum-Tarski algebra $\mathcal{LT}$. Of course,

$$v^*(A) = [A].$$

We define a mapping $v_0 : VAR \rightarrow \{1, 0\}$ as a composition of v and h , that is, we put, for any $a \in VAR$, $v_0(a) = hv(a) = h([a])$. The extension v_0^* of v_0 defined as $v_0^*(A) = hv^*(A)$ is a homomorphism of the algebra $\mathcal{F}$ of formulas and the two-element Boolean algebra and

$$v_0^*(A) = h([A]) = [[A]]_{\Delta} = 0.$$

This proves that v_0 is a **counter-model** for A , that is,

$$\models \neq A,$$

what **ends** the proof.

Here is a full statement of algebraic generalization of the Gödel completeness for predicate classical logic as it was first formulated and proved in Rasiowa and Sikorski (1950).

Theorem 2 (Rasiowa Sikorski 1950)

For every formula A of the classical predicate calculus $S = \{\mathcal{L}, \mathcal{C}\}$ the following conditions are equivalent:

- (i) A is derivable in S .
- (ii) A is valid in every realization of $\mathcal{L}$.
- (iii) A is valid in every realization of $\mathcal{L}$ in any complete Boolean algebra.
- (iv) A is valid in every realization of $\mathcal{L}$ in the field $B(X)$ of all subsets of any set $X \neq \emptyset$.
- (v) A is valid in every semantic realization of $\mathcal{L}$ in any enumerable set.
- (vi) There exists a nondegenerate Boolean algebra $\mathcal{A}$ and an infinite set J such that A is valid in every realization of $\mathcal{L}$ in J and $\mathcal{A}$.

- (vii) $A_R(\mathbf{i}) = 1$ for the canonical realization R of $\mathcal{L}$ in the Lindenbaum-Tarski algebra $\mathcal{LT}$ of S and the identity valuation $\mathbf{i}$.
- (viii) A is a predicate tautology.

Example: Some Non-Boolean Algebras and Nonclassical Logics

E1. Positive Implication Algebras

The positive implication algebras (PIA) are algebraic models for the positive implication logic (PIL).

The positive implication logic was developed in Hilbert and Bernays (1934) in search of axiomatization for purely implicative part of the intuitionistic logic (IL). This was proved almost 30 years later by Horn (1962) in the form of the following Separation Theorem.

For any formula A of the language with implication as the only connective, A is provable in PIL if and only if A is provable in IL.

The Separation Theorem fails (Horn 1962) for the corresponding predicate logics. The predicate calculus of PIL was examined for the first time from the algebraic point of view in Henkin (1950) under the name of *implicative models*. The implicative models are dual to the positive implication algebras presented in Rasiowa (1974), as is the theory of *Hilbert algebras* developed in Diego (1965). We follow here the definitions and formalization presented in Rasiowa (1974).

An abstract algebra $(A, 1, \Rightarrow)$ with a 0-argument operation 1 and a two-argument operation $\Rightarrow$ is called a **positive implication algebra (PIA)** if the following conditions hold:

1. $a \Rightarrow (b \Rightarrow a) = V$
2. $(a \Rightarrow (b \Rightarrow c)) \Rightarrow ((a \Rightarrow b) \Rightarrow (a \Rightarrow c)) = V$
3. if $a \Rightarrow b = V$ and $b \Rightarrow a = V$, then $a = b$
4. $a \Rightarrow V = V$

The class of all positive implication algebras is equationally definable (?), that is, the following theorem holds, as proved in Rasiowa (1974).

Equational Definability of PIA In every PIA algebra $(A, 1, \Rightarrow)$ the following equations hold:

- I1** $(a \Rightarrow a) \Rightarrow b = b$,
- I2** $a \Rightarrow (b \Rightarrow c) = (a \Rightarrow b) \Rightarrow (a \Rightarrow c)$,
- I3** $(a \Rightarrow b) \Rightarrow ((b \Rightarrow a) \Rightarrow b) = (b \Rightarrow a) \Rightarrow ((a \Rightarrow b) \Rightarrow a)$.

Moreover, if $(A, \Rightarrow)$ is an algebra with a two-argument operation $\Rightarrow$ such that for all $a, b, c \in A$ the equations **I1–I3** are satisfied, then for all $a, b \in A$ $a \Rightarrow a = b \Rightarrow b$, and $(A, 1, \Rightarrow)$ is a PIA algebra.

Positive Implication Algebra of Sets Let X be a topological space with an interior operation **I**. Denote by $G(X)$ the class of all open subsets of X . Then $(G(X), X, \Rightarrow)$, where $Y \Rightarrow Z = \mathbf{I}((X - Z) \cup Z)$ for all $Y, Z \in G(X)$, is a **PIA algebra**,

Any subalgebra of the PIA algebra $(G(X), X, \Rightarrow)$ defined above is called a PIA algebra of sets.

Theorem 3 (Representation Theorem for PIA)

Every PIA algebra is isomorphic to a PIA algebra of sets. More precisely, for every PIA algebra $\mathcal{A} = (A, 1, \Rightarrow)$ there exists a monomorphism h of $\mathcal{A}$ into a PIA algebra $(G(X), X, \Rightarrow)$ of all open subsets of a topological T_0 -space X .

The Representation Theorem was first proved in Diego (1965). The above version comes from Rasiowa (1974).

E2. Implication Algebras

The implication algebras (IA) are **algebraic models** for the classical implicative logic (CIL). The classical implicative logic CIL was developed by Tarski and Łukasiewicz (Łukasiewicz 1929) as the purely implicative part of classical logic (CL). The following theorem was proved by Tarski.

Theorem 4 (Separation Theorem)

For any formula A of the language with implication as the only connective, A is provable in IL if and only if A is provable in the classical logic CL.

The IA are PIA with an additional property. The theory of IA algebras has been developed in Abbot (1967).

Implication Algebra IA An abstract algebra $(A, 1, \Rightarrow)$ with a zero-argument operation 1 and a two-argument operation $\Rightarrow$ is called an **implication algebra IA** if it is a positive implication algebra PIA and in addition the following equation holds for any $a, b \in A$:

$$\mathbf{P} \quad (a \Rightarrow b) \Rightarrow a = a.$$

From the above definition and the Equational Definability Theorem for PIA we deduce the class of all IA is also equationally definable, that is, that the following is true.

Equational Definability of IA In every IA algebra $(A, 1, \Rightarrow)$ the following equations hold:

- I1** $(a \Rightarrow a) \Rightarrow b = b,$
- I2** $a \Rightarrow (b \Rightarrow c) = (a \Rightarrow b) \Rightarrow (a \Rightarrow c),$
- I3** $(a \Rightarrow b) \Rightarrow ((b \Rightarrow a) \Rightarrow b) = (b \Rightarrow a) \Rightarrow ((a \Rightarrow b) \Rightarrow a),$
- P** $(a \Rightarrow b) \Rightarrow a = a.$

On the other hand, if in an abstract algebra $(A, \Rightarrow)$ the conditions **I1–I3, P** hold, then for all $a, b \in A$ $a \Rightarrow a = b \Rightarrow b$, and $(A, 1, \Rightarrow)$ is an IA algebra.

Implication Algebra of Sets Let $B(X)$ be the class of all subsets of $X \neq \emptyset$. We define, for all $Y, Z \in B(X)$

$$Y \Rightarrow Z = (X - Y) \cup Z$$

and call every subalgebra of $(B(X), X, \Rightarrow)$ an implication algebra of sets.

The proof of the representation theorem for IA algebras is a slight modification of the proof of similar theorem for PIA algebras, and so is the theorem itself.

Theorem 5 (Representation Theorem for IA)

For every IA algebra $\mathcal{A} = (A, V, \Rightarrow)$ there exists a monomorphism h of $\mathcal{A}$ into an IA algebra $G(X)$, $(X, \Rightarrow)$ of all subsets of a space X . Hence, every IA algebra is isomorphic to an IA algebra of sets.

E3. Peirce Algebras

Peirce algebras (PA) were introduced in Rasiowa (1974) book *An Algebraic Approach to Non-Classical Logics* in the Exercises section (page 48). Almost all exercises presented in the book, despite their unassuming name, are previously published results, with proper references, or they are open problems. The questions asked about the Peirce algebras were the following:

- Q1. Prove that every Peirce algebra is a distributive lattice with the unit element 1 .
- Q2. Prove that the notion of an implicative filter Δ in PA algebra coincides with the notion of a filter.
- Q3. Let $\mathcal{A}$ be a PA algebra, prove that the quotient algebra $\mathcal{A}/\nabla$ is a PA algebra.
- Q4. Prove the Representation Theorem for the PA algebras.

Professor Helena Rasiowa gave the author in a private communication (June 1993) the following suggestions and information:

- 1. Peirce Algebras were defined in order to provide an algebraic model for a negation-free fragment of classical logic. They got their name from the fact that the Peirce Law $((a \Rightarrow b) \Rightarrow a) \Rightarrow a$ holds in the corresponding logic. Let's denote this logic PL. Then one can prove (using Q3.) as the essential step the following Separation Theorem for PL logic.

Let A be a negation free propositional formula, then A is provable in classical logic CL if and only if A is provable in PL logic.

- 2. One can use the implicative filters (Q2) in order to prove the Representation Theorem (Q4).
- 3. The solution to the Q1 becomes a simple corollary to the Representation Theorem.
- 4. The direct proof, not via the Representation Theorem, of the **Problem 1** is not known.

We present here (as an example of the complexity of mathematical reasoning involved) the proof of the **Representation Theorem** and hence an algebraic proof of the Q1.

Peirce Algebra PA By a Peirce algebra (PA) we mean any algebra

$$\mathcal{A} = (A, 1, \Rightarrow, \cup, \cap),$$

where $(A, 1, \Rightarrow)$ is an implication algebra IA and the following conditions are satisfied for all $a, b, c, \in A$:

- P1** $a \cup b = (a \Rightarrow b) \Rightarrow b$,
- P2** $(a \cap b) \Rightarrow a = 1$,
- P3** $(a \cap b) \Rightarrow b = 1$,
- P4** $(a \Rightarrow b) \Rightarrow ((a \Rightarrow c) \Rightarrow (a \Rightarrow (b \cap c))) = 1$.

Obviously, the class of all PA algebras is equationally definable.

Equational Definability of PA In every PA algebra $(A, \Rightarrow, \cup, \cap)$ the following equations hold:

- I1** $(a \Rightarrow a) \Rightarrow b = b$,
- I2** $a \Rightarrow (b \Rightarrow c) = (a \Rightarrow b) \Rightarrow (a \Rightarrow c)$,
- I3** $(a \Rightarrow b) \Rightarrow ((b \Rightarrow a) \Rightarrow b) = (b \Rightarrow a) \Rightarrow ((a \Rightarrow b) \Rightarrow a)$,
- P** $(a \Rightarrow b) \Rightarrow a = a$,
- P1** $a \cup b = (a \Rightarrow b) \Rightarrow b$,
- P2** $(a \cap b) \Rightarrow a = 1$,
- P3** $(a \cap b) \Rightarrow b = 1$,
- P4** $(a \Rightarrow b) \Rightarrow ((a \Rightarrow c) \Rightarrow (a \Rightarrow (b \cap c))) = 1$.

On the other hand, if in an abstract algebra $(A, \Rightarrow)$ the conditions **I1–I3**, **P**, **P1–P4** hold, then for all $a, b \in A$ $a \Rightarrow a = b \Rightarrow b$, and $(A, 1, \Rightarrow, \cup, \cap)$ is a PA algebra.

Peirce Algebra of Sets Let $B(X)$ be the class of all subsets of $X \neq \emptyset$. Then

$$(B(X), X, \Rightarrow, \cup, \cap),$$

where $\cup, \cap$ are set theoretical union and intersection, and for any $Y, Z \in B(X)$ we define

$$Y \Rightarrow Z = (X - Y) \cup Z$$

is a PA algebra and any of its subalgebras is called a PA algebra of sets.

Now we can formulate the following:

Theorem 6 (Representation Theorem for PA)

Every PA algebra is isomorphic to a PA algebra of sets. More precisely, for every PA algebra $\mathcal{A}$ there exists a monomorphism h of $\mathcal{A}$ into a PA algebra $(G(X), X, \Rightarrow, \cup, \cap)$ of all subsets of a space X .

By the definition any PA algebra of sets is a distributive lattice, so must be any PA algebra isomorphic with its proper PA algebra of sets. Hence we get the solution of the **Problem 1** as a corollary of the Representation Theorem.

In order to prove the Representation Theorem for PA we have to introduce some additional notions.

Filters in Peirce Algebras

Implicative Filter A subset $\nabla \subset A$ of the set of all elements of a PA algebra $(A, 1, \Rightarrow, \cup, \cap)$ is called an implicative filter if the following conditions are satisfied.

- f1** $1 \in \nabla$
- f2** if $a \in \nabla$ and $a \Rightarrow b \in \nabla$, then $b \in \nabla$.

Prime Implicative Filter A proper implicative filter ∇ is called prime if for all $a, b, \in A$ the following condition holds.

- f** If $a \cup b = (a \Rightarrow b) \Rightarrow b \in \nabla$, then either $a \in \nabla$ or $b \in \nabla$.

Irreducible Implicative Filter A proper implicative filter ∇ is called irreducible if for any two proper implicative filters ∇_1, ∇_2 such that $\nabla = \nabla_1 \cap \nabla_2$, either $\nabla = \nabla_1$ or $\nabla = \nabla_2$. I. e. a proper implicative filter is irreducible if it is not the intersection of two proper implicative filters different from it.

The following facts are essential to the proof of the Representation Theorem. The proofs are exactly the same as in Rasiowa (1974).

- F1** If in a PA algebra $(A, \Rightarrow, \cup, \cap)$ $a \leq b$ does not hold for some $a, b, \in A$, then there exists an irreducible implicative filter $\nabla \subset A$ such that $a \in \nabla$ and $b \notin \nabla$.

F2 An implicative filter in PA algebra is prime if and only if it is irreducible.

Proof of the Representation Theorem

Let X be the class of all prime filters in a PA algebra $\mathcal{A} = (A, 1, \Rightarrow, \cup, \cap)$. We define a mapping $h: A \rightarrow B(X)$ as follows:

$$h(a) = \{\nabla \in X : a \in \nabla\}, \quad (6)$$

for all $a \in A$. We have to show that the mapping h satisfies the conditions:

- c1** h is one-to-one,
- c2** $h(a \Rightarrow b) = h(a) \Rightarrow h(b)$,
- c3** $h(a \cup b) = h(a) \cup h(b)$,
- c4** $h(a \cap b) = h(a) \cap h(b)$,
- c5** $h(1) = X$.

Proof of c1: Let $\leq$ be the order relation defined in **F1**. We prove first that

$$a \leq b \quad \text{if and only if} \quad h(a) \subseteq h(b), \text{ for } a, b \in A. \quad (7)$$

Observe that $a \leq b$ is equivalent to $a \Rightarrow b = 1$. If $\nabla \in h(a)$ then $a \in \nabla$. On the other hand $a \Rightarrow b = 1 \in \nabla$, hence by **f2** of the definition, $b \in \nabla$. Consequently, $\nabla \in h(b)$ and $h(a) \subseteq h(b)$.

Assume now that $a \not\leq b$. Then by the **F1** there is an irreducible implicative filter ∇ such that $a \in \nabla$ and $b \notin \nabla$, that is, $\nabla \in h(a)$ and this ends the proof of (7).

If $a \neq b$, then one of the conditions $a \leq b$, $b \leq a$ does not hold and by (7) one of the inclusions $h(a) \subseteq h(b)$, $h(b) \subseteq h(a)$ does not hold as well. Hence $h(a) \neq h(b)$, that is, h is one-to-one.

Proof of c2: By **f1** of the Definition of the Implicative Filter, $V \in \nabla$ for every $\nabla \in X$; consequently, by (6) every ∇ in X belongs to $h(1)$.

Proof of c3: Observe first that the condition **f2** of the implicative filter is logically equivalent to the fact that $a \Rightarrow b \in \nabla$ implies that $a \notin \nabla$ or $b \in \nabla$. Thus $\nabla \in h(a \Rightarrow b)$ implies that $\nabla \in (X - h(a)) \cup h(b)$. Hence

$$h(a \Rightarrow b) \subseteq (X - h(a)) \cup h(b). \quad (8)$$

It remains to prove that

$$(X - h(a)) \cup h(b) \subseteq h(a \Rightarrow b). \quad (9)$$

Suppose that $\nabla \in (X - h(a)) \cup h(b)$. If $\nabla \in h(b)$, then $b \in \nabla$. By the condition 1 of the PIA definition $b \Rightarrow (a \Rightarrow b) = 1 \in \nabla$ and by **f1**, **f2** of the implicative filter definition, we infer that $(a \Rightarrow b) \in \nabla$, that is, $\nabla \in h(a \Rightarrow b)$. If $\nabla \in (X - h(a))$, that is, $a \notin \nabla$ and $(a \Rightarrow b) \in \nabla$. Observe that by **P1**, **P3** of equational definability, $a \cup (a \Rightarrow b) = (a \Rightarrow (a \Rightarrow b)) \Rightarrow (a \Rightarrow b) = 1 \in \nabla$. But by **F2** ∇ is a prime filter, hence either $a \in \nabla$ or $a \Rightarrow b \in \nabla$. Since $a \notin \nabla$ we get $a \Rightarrow b \in \nabla$ and this ends the proof of **c3**.

Proof of c4: By **P1** of the of equational definability, $a \cup b = (a \Rightarrow b) \Rightarrow b$. This and **c3** gives that $h(a \cup b) = h((a \Rightarrow b) \Rightarrow b) = (h(a) \Rightarrow h(b)) \Rightarrow h(b) = h(a) \cup h(b)$.

Proof of c5: By (6) **c5** is equivalent to

$$a \cap b \in \nabla \text{ if and only if } a \in \nabla \text{ and } b \in \nabla.$$

Assume that $a \cap b \in \nabla$. Directly from **P2**, **P3** of equational definability and the implicative filter definition, we get $(a \cap b) \Rightarrow a$, $(a \cap b) \Rightarrow b \in \nabla$, so $a \in \nabla$ and $b \in \nabla$.

Assume now that $a \in \nabla$ and $b \in \nabla$. By property 1 of the definition of PIA, $b \Rightarrow (a \Rightarrow b) = 1 \in \nabla$, but $b \in \nabla$, hence by definition of the implicative filter, $(a \Rightarrow b) \in \nabla$. By **P4** of equational definability $(a \Rightarrow a) \Rightarrow ((a \Rightarrow b) \Rightarrow (a \Rightarrow (a \cap b))) = 1 \in \nabla$ and $a \Rightarrow a = V \in \nabla$, $(a \Rightarrow b) \in \nabla$, hence by definition of the implicative filter, $a \cap b \in \nabla$.

This ends the proof of the Representation Theorem.

E4. Quasi-Boolean Algebras and De Morgan Lattices

De Morgan lattices have been introduced and examined in Moisil (1935). They are distributive lattices with an additional operation $\sim$, defined by two axioms: $\sim \sim a = a$, $\sim(a \cup b) = \sim a \cap \sim b$.

The quasi-Boolean algebras were introduced and examined in Białynicki-Birula and Rasiowa (1957). They are de Morgan lattices with a unit

element and a zero element. They are clearly a slight generalization of Boolean algebras. The theory of the quasi-Boolean algebras applies to the theory of quasi-pseudo-Boolean algebras that are models for a constructive logic with strong negation (Nelson 1949; Markov 1950; Vorobiev 1952; Thomason 1969).

Quasi-Boolean Algebra An abstract algebra

$$(A, 1, \cup, \cap, \sim) \quad (10)$$

is a **quasi-Boolean algebra (QBA)** if for all $a, b, c \in A$ the following equations are satisfied.

- I1** $a \cup b = b \cup a, a \cap b = b \cap a,$
- I2** $a \cup (b \cap c) = (a \cup b) \cap c, (a \cap b) \cup c = a \cap (b \cup c),$
- I3** $(a \cap b) \cup b = b, a \cap (a \cup b) = a,$
- I4** $a \cap (b \cup c) = (a \cap b) \cup (a \cap c), a \cup (b \cap c) = (a \cup b) \cap (a \cup c),$
- I5** $a \cup 1 = 1, a \cap 1 = a,$
- q1** $\sim \sim a = a,$
- q2** $\sim(a \cup b) = \sim a \cap \sim b.$

De Morgan Lattice An abstract algebra

$$(A, 1, \cup, \cap, \sim) \quad (11)$$

is a de Morgan lattice (DML) if the conditions **I1–I4, q1, q2** of the above definition of QBA algebra are satisfied.

QBA of Sets Let $X \neq \emptyset$ and let g be an involution of X , that is, $g: X \rightarrow X$ such that for all $x \in X$, $g(g(x)) = x$. We put, for any $Y \subseteq X$,

$$\sim Y = X - g(Y).$$

Let $\mathcal{Q}(X)$ be a non-empty class of subsets of X , containing X and closed under set-theoretical union $\cup$, intersection $\cap$, and the operation $\sim$ defined above. The algebra

$$(\mathcal{Q}(X), X, \cup, \cap, \sim)$$

is called a QBA algebra of sets, or a **quasi-field** of subsets of X .

Theorem 7 (Representation Theorem for QBA)

Every QBA algebra is isomorphic to a quasi-field of subsets of certain open subsets of a topological, compact T_0 space.

Granular Computing

The idea of rough sets is a mathematical concept introduced in Pawlak (1982). It was shown in Pawlak (1985b) that the concept of a rough set is different from that of a fuzzy set. Further relationship between those two concepts was studied in Dubois and Prade (1987). It was demonstrated in Wong and Ziarko (1987) that the generalized notion of rough sets based on the probabilistic approximation space can be conveniently described by the concept of fuzzy sets. Usually rough set approaches used at that time purely semantical methods in Ras and Zemankova (1986) where mixed semantical and syntactical (formal system) methods were used. The purely syntactical methods and problems that can be solved by them within the framework of rough set investigations are presented in Wasilewska (1989) and Hadjimichael and Wasilewska (1993).

The automated inference of classification rules (then called decision rules in the Rough Sets based investigations) was the subject of the works of many authors starting with Michalski's AQ11 in Michalski and Larson (1978) and Quinlan's ID3 in Quinlan (1983) algorithms and other Decision Trees algorithms. The problem was that usually they produced an unwieldy number of rules. Rough Sets based classification models and algorithms were invented and proved that they were able to reduce the number of rules without losing their accuracy. They were additionally extended to a more general nondeterministic model by adding probabilistic approximations in Wong et al. (1986).

I was proved in Orłowska (1983) that propositional aspects of rough set theory are adequately captured by the modal system S5. In this case a Kripke model gives the approximation space

$(A, \emptyset, R)$ in which the well-formed formulas are interpreted as *rough sets*.

Following Orłowska results, Banerjee introduced in Banerjee and Chakraborty (1993) a new binary connective $\sim$ in the modal logic S5, the intended interpretation of which is the notion of the *rough equality*. She then used a construction similar to the construction of Lindenbaum-Tarski algebra (5) on the set of all formulas of the logic S5 with additional connective $\sim$. The resulting algebra corresponding to the equivalence relation defined by new connective $\sim$ is called an S5 Rough Algebra (19).

The work in Banerjee (1992) and Banerjee and Chakraborty (1993) includes results of an extensive syntactic study of properties of the S5 Rough Algebra that are provable within the modal logic S5. The open question of providing a *semantic component* for it was addressed and presented in Wasilewska and Banerjee (1995).

These works consequently led to discovery and study of classes algebraic models inspired by Rough Sets investigations (Wasilewska 1996, 1997; Wasilewska and Vigneron 1997, 1998).

We present here these results which by the use algebraic logic techniques give a deeper mathematical meaning to the rough sets theory in general, and to the notion of rough equality, introduced by Pawlak (1982), in particular.

We first present short overview of the work in Banerjee (1992) and Banerjee and Chakraborty (1993) leading to the definition of *S5 Rough Algebra*. Then we show that the S5 Rough Algebra is a particular case of a Quasi-Boolean algebra discussed in section “[Example: Some Non-Boolean Algebras and Non-Classical Logics](#).” In the next step we introduce, justify, and discuss two new, rough sets inspired classes of algebras: Topological Quasi-Boolean algebras and Topological Rough algebras.

We also show that Białynicki-Birula and Rasiowa’s proof of a representation theorem for quasi-Boolean algebras (Białynicki-Birula and Rasiowa 1957) can be generalized, via MacKinsey and Tarski’s proof for the topological Boolean algebras (McKinsey and Tarski 1948), to the proof of the Representation Theorems for these new classes of algebras.

S5 Rough Algebra

To make this section self-contained we review now some basic definitions and explain the main points of the definition of the S5 Rough Algebra.

Approximation space

Let U be a non-empty set called a universe, and let R be an equivalence relation on U . The triple $(U, \emptyset, R)$ is called an approximation space.

Lower, upper approximations

Let $(U, \emptyset, R)$ and $A \subseteq U$.

Denote by $[u]$ the equivalence class of R . The sets

$$\underline{A} = \cup\{[u] \in A/R : [u] \subseteq A\}, \quad \overline{A} = \cup\{[u] \in A/R : [u] \cap A \neq \emptyset\}$$

are called *lower* and *upper* approximations of A , respectively.

Rough equality

Given an approximation space $(U, \emptyset, R)$ and any $A, B \subseteq U$. We say that the sets A and B are *roughly equal* and denote it by $A \sim_R B$ if and only if $\underline{A} = \underline{B}$ and $\overline{A} = \overline{B}$.

Boolean algebra

An abstract algebra $(A, 1, 0, \Rightarrow, \cap, \cup, \neg)$ is said to be a Boolean algebra if it is a distributive lattice and every element $a \in A$ has a complement $\neg a \in A$.

Topological Boolean algebra

By a topological Boolean algebra we mean an abstract algebra $(A, 1, 0, \Rightarrow, \cap, \cup, \neg, I)$ where $(A, 1, 0, \Rightarrow, \cap, \cup, \neg)$ is a Boolean algebra and, moreover, the following conditions hold: $I(a \cap b) = Ia \cap Ib$, $Ia \cap a = Ia$, $I Ia = Ia$, and $I1 = 1$, for any $a, b \in A$.

The element Ia is called a *interior of a* . The element $\neg I \neg a \neg I \neg a$ is called a *closure of a* and will be denoted by Ca . Thus the operations I and

C are such that $Ca = \neg I \neg a$ and $Ia = \neg C \neg a$. The element a is said to be *open* (*closed*) if $a = Ia$ ($a = Ca$).

Before describing their construction leading to the definition of the S5 Rough Algebra, we include below a description of a standard construction of a Lindenbaum-Tarski algebra for the modal logic S5.

Construction of Modal Lindenbaum-Tarski Algebra

Given a modal propositional language $\mathcal{L}$ with of connectives $\{\cup, \cap, \Rightarrow, \neg, \square, \diamond\}$ and the set F of formulas.

We form the algebra of formulas

$$\mathcal{F} = (F, \cup, \cap, \Rightarrow, \neg, \square, \diamond), \quad (12)$$

whereby performing an operation on a formula (two formulas) means writing the formula having this operation as a main connective. For example

$$\begin{aligned} \cup(A, B) &= (A \cup B), \quad \Rightarrow(A, B) \\ &= (A \Rightarrow B), \quad \square(A) = \square A, \quad \diamond(A) = \diamond A. \end{aligned}$$

Let $S = (\mathcal{L}, Cn)$ be a modal deductive system (logic) with a consequence operation Cn given by a certain set of logical axioms AX and some rules of inference. We write

$$\vdash_S A$$

to denote that a formula A has a formal proof in S .

We define two binary relations $\leq$ and $\approx$ in the algebra $\mathcal{F}$ of formulas (12) as follows. For any $A, B \in F$,

$$A \leq B \text{ if and only if } \vdash_S (A \Rightarrow B), \quad (13)$$

$$A \approx B \text{ if and only if } \vdash_S (A \Rightarrow B) \text{ and } \vdash_S (B \Rightarrow A). \quad (14)$$

In order to construct a modal Lindenbaum-Tarski algebra for a deductive system (logic) $S = (\mathcal{L}, Cn)$ its sets of axioms and rules of inference must be such that it is possible to prove all facts listed below:

The relation $\leq$ is a quasi-ordering in F .

The relation $\approx$ is an equivalence relation in F . We denote the equivalence class containing a formula A by $[A]$.

The quasi ordering $\leq$ on the set of formulas F induces an ordering relation on $F/\approx$ defined as follows: $[A] \leq [B]$ if and only if $A \leq B$.

The equivalence relation $\approx$ on F is a congruence with respect to all logical connectives $\{\cup, \cap, \Rightarrow, \neg, \square, \diamond\}$.

Modal Lindenbaum-Tarski Algebra

An algebra

$$\mathcal{M} = (F/\approx, \cup, \cap, \Rightarrow, \neg, \square, \diamond), \quad (15)$$

such that its operations $\cup, \cap, \Rightarrow, \neg, \square$, and $\diamond$ are determined by the congruence relation (14), that is,

$$\begin{aligned} [A] \cup [B] &= [(A \cup B)], \quad [A] \cap [B] \\ &= [(A \cap B)], \quad [A] \Rightarrow [B] = [(A \Rightarrow B)], \end{aligned}$$

$$\begin{aligned} \neg[A] &= [\neg A], \quad \square A = [\square A], \quad \text{and} \quad CA \\ &= [\diamond A]. \end{aligned}$$

is called a modal Lindenbaum-Tarski algebra.

In the case of modal logic S4 the Modal Lindenbaum-Tarski algebra (see McKinsey 1941; McKinsey and Tarski 1948; Rasiowa and Sikorski 1953) is a topological Boolean algebra.

In the case of modal logic S5 it is topological Boolean algebra such that every open element is closed and every closed element is open.

Moreover, in both cases, the following holds.

Fact 3.1 For any formula A , $\vdash_{S4} A$ if and only if $[A] = 1$ and $\vdash_{S5} A$ if and only if $[A] = 1$.

S5 Rough Algebra Construction

We add now a new binary connective $\sim$ in the language of the modal logic S5. The intended interpretation of it is the notion of the *rough equality*, hence we call it a rough equality connective. It is defined in terms of standard S5 connectives $\{\cup, \cap, \rightarrow, \Leftrightarrow, \neg, \Box, \Diamond\}$ as follows.

Rough equality connective $\sim$

For any formulas A, B (of the modal S5 language), we write $A \sim B$ for the formula $((\Box A \Leftrightarrow \Box B) \cap (\Diamond A \Leftrightarrow \Diamond B))$.

In the next step we use the connective $\sim$ and a construction similar to that of the Modal Lindenbaum-Tarski algebra (15) but based on the set of all formulas of S5 with additional connective $\sim$. We define now a new binary relation $\approx$ on the set F of formulas of the modal S5 logic as follows. For any $A, B \in F$,

$$A \approx B \text{ if and only if } \vdash_{S5} A \sim B. \quad (16)$$

that is, we define

$$A \approx B \text{ if and only if } \vdash_{S5} ((\Box A \Leftrightarrow \Box B) \cap (\Diamond A \Leftrightarrow \Diamond B)).$$

We prove that the above relation (16) corresponding to the notion of rough equality is an *equivalence relation* on the set F of formulas of S5.

We define binary relations $\leq$ on $F/\approx$ as follows:

$$[A] \leq [B] \text{ if and only if } \vdash_{S5} ((\Box A \Rightarrow \Box B) \cap (\Diamond A \Rightarrow \Diamond B)).$$

We prove that $\leq$ is an order relation on $F/\approx$ with the greatest element $1 = [A]$, for any formula A , such that $\vdash_{S5} A$, and with the least element $0 = [B]$, such that $\vdash_{S5} \neg B$.

We prove that $\approx$ is a congruence relation with respect to $\neg, \Box, \Diamond$.

We prove that $\approx$ is **not** a congruence relation with respect to $\Rightarrow, \cap$, and $\cup$.

We introduce two new operations $\sqcup$ and $\sqcap$ in $F/\approx$ as follows:

$$[A] \sqcup [B] = [(A \cup B) \cap (A \cup \Box A \cup \Box B \cup \neg \Box(A \cup B))], \quad (17)$$

$$[A] \sqcap [B] = [(A \cap B) \cup (A \cap \Diamond A \cap \Diamond B \cap \neg \Diamond(A \cap B))]. \quad (18)$$

Definition 1 (S5 Rough Algebra)

An abstract algebra

$$\mathcal{R} = (F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1), \quad (19)$$

such that the operations $\sqcup$ and $\sqcap$ are defined by (17) and (18), respectively, the operations $\neg, I$, and C are induced, as in the S5 Lindenbaum-Tarski algebra, that is, by the relation $\approx$ defined, as

$A \approx B$ if and only if $\vdash_{S5} (A \Rightarrow B)$ and $\vdash_{S5} (B \Rightarrow A)$ is called an S5 rough algebra.

Properties of the S5 Rough Algebra

In Banerjee (1992) many important properties of the rough algebra were proved. They were also reported in Banerjee and Chakraborty (1993). We cite here only those that are relevant to our future investigations.

- P1** $(F/\approx, \leq, \sqcup, \sqcap, 0, 1)$ is a distributive lattice with 0 and 1.
- P2** For any $[A], [B] \in F/\approx$, $\neg([A] \sqcup [B]) = (\neg[A] \sqcap \neg[B])$.
- P3** For any $[A] \in F/\approx$, $\neg \neg [A] = [A]$.
- P4** The rough algebra is not a Boolean algebra, that is, there is a formula A of a modal logic S5, such that $\neg[A] \sqcap [A] \neq 0$ and $\neg[A] \sqcup [A] \neq 1$.
- P5** For any $[A], [B] \in F/\approx$, $I([A] \sqcap [B]) = (I[A] \sqcap I[B])$, $I[A] \leq [A]$, $II[A] = I[A]$, $II = 1$, and $CI[A] = I[A]$, where $C[A] = \neg I \neg [A]$.

Questions and Observations

The above and other properties of the rough algebra proved in Banerjee (1992) lead to some natural questions and observations.

- Q1** By the property **P4**, the S5 Rough algebra's complement operation $\neg$ is not a Boolean complement. Let's call it a *rough complement*. We can see that the rough complement is pretty close to the Boolean

complement because the other de Morgan law

$$\neg([A] \sqcap [B]) = (\neg[A] \sqcap \neg[B])$$

holds in the S5 Rough algebra.

Also he very Boolean laws

$$\neg 1 = 0 \quad \text{and} \quad \neg 0 = 1$$

hold in it, so we ask what kind of a complement is the rough complement.

- Q2** The S5 Rough algebra (19) is not, by P4, a Boolean algebra, so we ask which kind of algebra is it.
- Q3** Has such an algebra similar to the S5 Rough algebra been discovered and investigated before?

Observation 1 A complement operation with similar properties to the *rough complement* was introduced in 1935 by Moisil and leads to a definition of a notion of *de Morgan Lattices*. De Morgan lattices (11) are distributive lattices satisfying the conditions P2 and P3. They were investigated in Moisil (1935) and Henkin (1963).

Observation 2 In 1957 Białynicki-Birula and Rasiowa have used the de Morgan lattices to introduce a notion of a *quasi-Boolean algebra*.

The formal definition of the quasi-Boolean algebras (10) is the following.

Quasi-Boolean algebra (Białynicki-Birula and Rasiowa 1957) An abstract algebra $\mathcal{A} = (A, \sqcup, \sqcap, \sim, 1)$ is called a *quasi-Boolean algebra* if $(A, \sqcup, \sqcap, 1)$ is a distributive lattice with unit element 1 and for any $a, b \in A$, $\sim(a \sqcup b) = (\sim a \sqcap \sim b)$ and $\sim \sim a = a$.

One can easily prove that in every quasi-Boolean algebra the *zero* (0) element exists. From that and properties P1–P4, we get the following fact.

Fact 3.2 The S5 Rough algebra (19) is not a Boolean algebra, but is a quasi-Boolean algebra.

Observation 3 The property P5 tells that the operations I and C of the S5 Rough algebra $\mathcal{R} = (F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1)$ fulfill the axioms of the topological Boolean algebra.

Definition 2 (Topological Quasi-Boolean Algebra)

An algebra $(A, \sqcap, \sqcup, \sim, 1, I)$ is called a *topological quasi-Boolean algebra* if $(A, \sqcup, \sqcap, \sim, 1)$ is a quasi-Boolean algebra and for any $a, b \in A$, $I(a \sqcap b) = Ia \sqcap Ib$, $Ia \sqcap a = Ia$, $I Ia = Ia$, and $I 1 = 1$.

The element Ia is called a *quasi-interior* of a . The element $\sim I \sim a$ is called *quasi-closure* of a . It allows us to define in A a unary operation C such that $Ca = \sim I \sim a$. We can hence represent the topological quasi-Boolean algebra as an algebra $(A, \sqcap, \sqcup, \sim, I, C, 0, 1)$ similar to the rough algebra $(F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1)$. From P4 we immediately get the following.

Fact 3.3 The S5 rough algebra $\mathcal{R} = (F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1)$ is a topological quasi-Boolean algebra.

Topological Rough Algebras

The property P5 of the rough algebra tells us also that the operations I and C fulfill an additional property: for any $[A] \in F/\approx$, $CI[A] = I[A]$. This justifies the following definition.

Definition 3 (Topological Rough Algebra)

A topological quasi-Boolean algebra $(A, \sqcap, \sqcup, \sim, I, C, 0, 1)$ such that for any $a \in A$, $CIa = Ia$, is called a *topological rough algebra*.

Note that the class of all topological quasi-Boolean algebras, and the class of all topological rough algebras are *equationally definable*.

Directly from above we get the following answer to the question Q3.

Fact 4.1

The S5 Rough algebra $\mathcal{R} = (F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1)$ is a topological rough algebra.

As we have said in the Introduction, one of the first questions one asks about a new algebra, or classes of algebras, is the existence and form of the representation theorem. This is a question about a deeper mathematical meaning of the newly created abstract algebras. We are going to introduce here all notions and steps that lead to the

understanding of meaning, complexity, and beauty of the representation theorem (and of its proof). We are not including the proof here. It is quite long and technical and will be published separately. Since the first question (**Q1**) we have asked here about the rough algebra was to provide some characterization of its complement operation, which we have called a *rough complement*, we will start with the mathematical characterization of this notion.

The Observation 1 and Fact 4.1 provided already some characterization of the rough complement, namely that it is a *quasi-Boolean complement*. We use and extend here results from Białynicki-Birula and Rasiowa (1957) to characterize further the notions of the rough complement, topological quasi-Boolean algebras, and topological rough algebras.

Let A be a non-empty set. We define after Białynicki-Birula and Rasiowa (1957) the following notion.

Involution Any mapping $g: A \rightarrow A$ such that for all $a \in A$, $g(g(a)) = a$ is called an *involution*. Clearly, every involution is a one-one mapping from A onto A .

Let $Q(A)$ be a non-empty class of subsets of A , containing A and closed under set-theoretical union and intersection, and under the operation $\sim$ defined as follows: for any $X \in Q(A)$, $\sim X = A - g(X)$. It is proved in Białynicki-Birula and Rasiowa (1957) that $(Q(A), A, \cup, \cap, \sim)$ is an example of a quasi-Boolean algebra. The quasi-Boolean algebra is a particular case (when the topology is given by the identity operation, i.e., when for any $a \in A$, $Ia = a$) of the topological quasi-Boolean algebra.

This justifies the following definition.

Definition 4 (Rough Complementation)

Given a non empty set A and an involution g on A . We call the operation $\sim X = A - g(X)$ a rough complementation of sets.

Definition 5 (Quasi-Field of Sets)

Given a non empty set A and the rough complementation $\sim$ in A , we call a structure $(Q(A), A, \cup, \cap, \sim)$ a quasi-field of subsets of A .

The representation theorem for the quasi-Boolean algebras says that quasi-fields of sets are typical examples of quasi-Boolean algebras (see Białynicki-Birula and Rasiowa 1957; Rasiowa 1974), that is, that the following holds.

Theorem 8 (Representation Theorem)

Every quasi-Boolean algebra is isomorphic to a quasi-field of certain open subsets of a topological, compact T_0 space.

We follow here basic notions of the algebraic logic, where the topological space is defined as follows.

Topological space

A set A is said to be a *topological space* if with every $X \subset A$ there is associated a set $IX \subset A$ such that the following conditions are satisfied: for any $X, Y \subset A$, $I(X \cap Y) = IX \cap IY$, $IX \subset X$, $IIX = X$, $IA = A$. The operation I is called the interior operation. The topological space is often written as (A, I) .

Given a topological space (A, I) , then $(Q(A), A, \cup, \cap, \sim, I)$ where $\sim$ is rough complement is a topological quasi-Boolean algebra. Clearly, every subalgebra of this algebra is also a topological quasi-Boolean algebra. This justifies the following definitions.

Definition 6 (Topological Quasi-Field of Sets)

Given a topological space (A, I) and a quasi-field of sets $(Q(A), A, \cup, \cap, \sim)$. We define the closure operation C as $CX = \sim I \sim X$ and call a structure $(Q(A), A, \cup, \cap, \sim, I, C)$ a topological quasi-field of sets or, more precisely, a topological quasi-field of subsets of A .

Definition 7 (Topological Rough-Field of Sets)

A topological quasi-field of sets $(Q(A), A, \cup, \cap, \sim, I, C)$ is called a topological rough-field of sets if additionally, for any set $X \in Q(A)$, $CIA = IA$.

The most important notion leading to the proof of the representation theorem for any algebra are the notions of a filter and ideal (see Stone 1934; Rasiowa 1974).

Filter A non-empty set ∇ of elements of universe A of an algebra with two binary operations $\cap$ and $\cup$ is said to be a *filter* in A provided that, for any elements $a, b \in A$, $a \cap b \in \nabla$ if and only if $a \in \nabla$ and $b \in \nabla$.

Ideal A non-empty set Δ of elements of A is said to be an *ideal* in A provided that, for any elements $a, b \in A$, $a \cup b \in \Delta$ if and only if $a \in \Delta$ and $b \in \Delta$.

A filter (ideal) is called *maximal* in A if it is proper and is not any proper subset of a proper filter (ideal) in A .

We adopt Rasiowa's definition of an I -filter (Rasiowa 1974) to our purposes, that is, we adopt the following definition.

Definition 8 (Rough Filter)

A filter ∇ (ideal Δ) in a topological quasi-Boolean algebra $(A, \cap, \cup, \sim, I, C, 0, 1)$ is called a *rough filter* (rough ideal) provided

$$a \in \nabla \text{ implies that } Ia \in \nabla \text{ for every } a \in A.$$

Then we show that all major properties of Rasiowa's I -filters hold for the rough-filters, in particular we show the following.

Fact 4.2 (Maximal Rough Filter)

For every element $a \neq 0$ in A there exists a maximal rough filter ∇ in A , such that $a \in \nabla$.

Given a topological quasi-Boolean algebra $(A, \cap, \cup, \sim, I, C, 0, 1)$, let us now put, for any set $S \subseteq A$,

$$\tilde{S} = \{\sim a \in A : a \in S\}.$$

We prove the following duality property.

Duality If ∇ is a rough maximal filter in a topological quasi-Boolean algebra, the set $\tilde{\Delta}$ is a rough maximal ideal.

Then we combine the techniques from Stone (1934), Rasiowa and Sikorski (1963), and from Białynicki-Birula and Rasiowa (1957) to prove that the following holds.

Theorem 9 (Representation Theorem)

For every topological quasi-Boolean algebra (topological rough algebra) $\mathcal{A}$ there exists a monomorphism h from, $\mathcal{A}$ into a topological quasi-field (rough-field) of sets of a topological space A .

Fact 4.3

The $S5$ Rough algebra $\mathcal{R} = (F/\approx, \sqcup, \sqcap, \neg, I, C, 0, 1)$ is isomorphic with a topological rough-field, of sets.

Generalized Rough Sets, L-Fuzzy, and LT-Fuzzy Sets

A concept of LT-fuzzy sets was introduced by Rasiowa and Cat Ho (1992). LT-fuzzy sets are a modification of L-fuzzy sets introduced by Goguen (1967). We introduce here a notion of a generalized rough set and show that it can be considered as a particular case of an L-fuzzy set. We also generalize the notion of rough equality of sets, introduced in Novotny and Pawlak (1985) to a notion of topological and rough topological equality of sets. We show that LT-fuzzy sets provide a common characterization for rough, generalized rough sets and topological and rough equality congruent spaces.

LT-Fuzzy Sets

A **fuzzy** set (Zadeh 1965) in a non-empty universe U is a set of ordered pairs $\{(f(u), u)\}$ for $u \in U$, where $f: U \rightarrow [0, 1]$ is a membership function and

$[0, 1]$ is a closed interval of a real line. We identify any fuzzy set with its membership function.

An **L-fuzzy** set (Goguen 1967) in a non-empty universe U is any function $f: U \rightarrow L$, where L is a poset, in particular a lattice, a complete lattice, an interval $[0, 1]$.

The **LT-fuzzy** sets (Rasiowa and Cat Ho 1992) in a non-empty universe U are functions assigning to objects in U certain subsets of a fixed **poset** $(T, \leq)$, interpreted as sets of membership values.

The motivation for introducing LT-fuzzy sets is the following:

In order to approximate a vague predicate such as “*beautiful x*” in a set U of cars, we use parameters that are not always linearly but often partially ordered with respect to their importance. These parameters and their partially ordered values are assumed as a **poset** $(T, \leq)$. Sets of the parameters’ values form LT-fuzzy membership values for objects in U .

The formal definition is the following:

Given a poset $(T, \leq)$, considered as a poset of membership values.

By a **T-fuzzy membership** value we mean any non-empty subset I of T such that for any $s, t \in T$ if $s \leq t$ and $t \in I$, then $s \in I$.

Let LT be a **family of subsets** of T consisting of the empty set and all T-fuzzy membership values.

An **LT-fuzzy set** in a non-empty universe U is any function

$$f : U \rightarrow LT.$$

Observe that for every poset T , LT is a poset and hence any LT-fuzzy set is an L-fuzzy set for $L = LT$.

Generalized Rough Sets

The idea of rough set was proposed in Pawlak (1982). The starting point of the rough set theory is the assumption that we have initially some information (knowledge) about the elements of the universe and that some elements can be indiscernible (equivalent) in the context of the information about them, that they can be “seen” as the same. These assumptions justified the following definitions (Pawlak 1991).

Approximation Space

Let U be a non-empty set called a universe, and let R be an equivalence relation on U . The tuple (U, R) is called an approximation space defined by the equivalence relation R . Let (U, R) and be the approximation space and $A \subseteq U$. Denote by $[u]$ the equivalence class of R . The sets

$$\begin{aligned}\underline{A} &= \cup\{[u] \in A/R : [u] \subseteq A\}, \\ \overline{A} &= \cup\{[u] \in A/R : [u] \subseteq A \neq \emptyset\}\end{aligned}$$

are called *lower* and *upper approximations* of A , respectively.

This means that any set A can be observed through its lower and upper approximations $\underline{A}$ and $\overline{A}$. A is observed precisely only if $\underline{A} = \overline{A}$. Otherwise A is observed only “roughly,” or more precisely, A is called a *rough set*.

We usually use a topological notation IA (interior) for the lower approximation $\underline{A}$ and CA (closure) for the upper approximation $\overline{A}$ because of their known (Pawlak 1991) topological properties.

Topological Space

We recall that a tuple (X, I) is called a topological space if for every $A \subseteq X$ there is associated a set $IA \subseteq X$, called the *interior of A*, in such a way that the following conditions are satisfied for any $A, B \subseteq X$:

$$\begin{aligned}I(A \cap B) &= IA \cap IB, \quad IA \subseteq A, \quad IIA \\ &= IA \text{ \textit{tex tan d} } IX = X.\end{aligned}$$

The operation I is then called the **interior operation**.

For every $A \subseteq X$, the set $-I - A$ is called the **closure of A** and denoted by CA . Thus by definition, $CA = -I - A$ and $IA = -C - A$.

Approximation Topological Space

By an approximation topological space we understand any topological space such that its interior operation is induced by an equivalence relation.

More exactly, let $(U, \approx)$ be an approximation space defined by an equivalence relation $\approx$). Let B be a class composed of $U, \emptyset$ and all $[u] \in U/\approx$.

It is easy to see that $\mathbf{B}$ is a basis of uniquely defined interior operation I , namely $IA = \bigcup \{B \in \mathbf{B} : B \subseteq A\}$ and the triple $(U, \approx, I)$ is called the *approximation topological space*. We define, as before, $CA = -I - A$. These definitions of the interior and closure operations in the case under consideration is equivalent with

$$IA = \bigcup \{[u] \in A / \approx : [u] \subset A\},$$

$$CA = \bigcup \{[u] \in A / \approx : [u] \cap A \neq \emptyset\}.$$

Moreover, it is easy to prove that the upper and lower approximations in the approximation space (U, R) and interior and closure operations in the topological approximation space $(U, \approx, I)$ have the following property. For any $A \subseteq U$,

$$CIA = IA \text{ and } ICA = CA.$$

Sets with such property are called *clopen sets* and the property is called a *clopen sets property*.

The following theorem was proven in Wiweger (1988).

Theorem 10 (Wiweger 1988)

Let X be a topological space with an interior operation I , having the clopen sets property. Then there exists an equivalence relation $\approx$ in X such that the interior operation I' , induced by $\approx$, coincides with I .

The above justifies the introduction of a following definition.

Definition 9 (Generalized Rough Set)

By a generalized rough set we will understand any topological space (X, I) with a clopen sets property.

Topological and Rough Equalities

The notion of *rough equality* was introduced Pawlak (1982) as follows. Given an approximation space (U, R) and any $A, B \subseteq U$. Let I, C denote the lower and upper approximation induced by R . We say that the sets A and B are *roughly equal* and denote it by $A \approx_R B$ if and only if $IA = IB$ and $CA = CB$. The rough equality was extensively examined in Novotny and Pawlak (1985), Pagliani

(1994), Banerjee and Chakraborty (1993), Vigneron and Wasilewska (1996), and others. It was examined in Banerjee and Chakraborty (1993) from the logical point of view and from algebraic point of view in Wasilewska (1997) and from algebraic-automated theorem proving angle in Wasilewska (1995). In all cases the notion of the rough equality was restricted to the rough sets only, that is, to the sets defined by the approximation spaces (U, R) and the lower and upper approximations induced by it. We propose here to extend it, in the natural way to the case of general topological spaces (X, I) , the approximate topological spaces $(X, \approx, I)$ and generalized rough sets as follows.

Definition 10

Given a topological space (X, I) , for any two sets $A, B \subset X$ we say that they are topologically equal, and denote

$$A \approx_T B$$

if and only if $IA = IB$ and $CA = CB$.

*In case when (X, I) is a clopen topological space, that is, in a case when the topological space (X, I) is a **generalized rough set** we say that the sets A, B are **topologically roughly equal** and denote*

$$A \approx_R B.$$

We leave the question of the detailed examination of the congruent spaces: $X/\approx_T$ and $X/\approx_R$ for the future investigations. Some preliminary work has already been done in Vigneron and Wasilewska (1996).

Rough, L-Fuzzy, and LT-Fuzzy Sets

The relationship between the rough and fuzzy sets has been discussed by many authors, in particular in Pawlak (1985b). We have established here the following relationship between the rough, generalized rough sets, topological equality, rough equality, and L-fuzzy and LT-fuzzy sets.

- R** Rough and generalized rough sets are particular cases of L-fuzzy sets.
- TE** The congruent space $X/\approx_T$ can be represented as a particular case of an L-fuzzy set.

- RE** The congruent space $X/\approx_R$ can be represented as a particular case of an LT-fuzzy set.
- LT** LT-fuzzy sets provide a common characterization for all above cases of rough, generalized rough sets, and topological and rough equality congruent spaces.

Bibliography

- Abbot JC (1967) Semi-Boolean algebras. *Mat Vesn* 4:177–198
- Banerjee M (1992) A categorical approach to the algebra and logic of the indiscernible. PhD dissertation, Mathematics Department, University of Calcutta
- Banerjee M, Chakraborty MK (1993) Rough consequence and rough algebra. In: Ziarko WP (ed) *Rough sets, fuzzy sets and knowledge discovery, of the international workshop on proceedings of Rough Sets and Knowledge Discovery, (RSKD'93), Banf, Alberta, Canada*. Springer, London, 1994, pp 196–207
- Beth EW (1951) A topological proof of the theorem of Löwenheim-Skolem-Gödel. In: Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam, Proceedings, Series A, vol 54, no. 5 and *Indagationes mathematicae*, vol 13, no. 5, pp 436–444
- Beth EW (1956) Semantic construction of intuitionistic logic. In: Koninklijke Nederlandse Akademie van Wetenschappen, Afd. Letterkunde, n.s. vol 19, no. 11, pp 357–388
- Beth EW (1959) *The foundations of mathematics*. North Holland Publishing Co., Amsterdam
- Białynicki-Birula A, Rasiowa H (1957) On the representation of quasi-Boolean algebras. *Bull Ac Pol Sci Cl III* 5:259–261
- Bronikowski Z (1994) Algebraic structures of rough sets. In: Ziarko W (ed) *Rough sets, fuzzy sets and knowledge discovery, workshops in computing*, Springer, pp 242–247
- Diego A (1965) Sobre algebras de Hilbert. *Notas de Logica Matematica*, Instituto de Matematica, Universidad Nacional del Sur, Bahia Bianca, p 12
- Dubois D, Prade H (1987) Twofold fuzzy sets and, rough, sets. *Fuzzy Sets Syst* 23:3–18
- Dyson VH, Kreisel G (1961) Analysis of Beth's semantic construction of intuitionistic logic. *Applied Mathematics and Statistics Laboratories*, Stanford University, Technical Report no.3, 27 Jan, pp 39–65
- Goguen JA (1967) L-fuzzy sets. *J Math Anal Appl* 18:145–174
- Hadjimichael M, Wasilewska A (1993) Interactive inductive learning. *Int J Man-Mach Stud* 38:147–167
- Halmos PR (1962) *Algebraic logic*. Chelsea, New York
- Henkin N (1950) An algebraic characterization of quantifiers. *Fundam Math* 37:63–74
- Henkin I (1955a) The representation theorem for cylindric algebras. *Studies in Logic and Foundations of Mathematics*, Amsterdam, pp 85–97
- Henkin I (1955b) Remarks on Henkin's paper: Boolean representation through propositional calculus. *Fundam Math* 41:89–96
- Henkin L (1963) A class of non-normal models for classical sentential logic. *J Symb Log* 28:300
- Henkin I, Tarski A (1961) Cylindric algebras. In: *Proceedings of Symposia in Pure Mathematics II*, pp 83–113
- Hilbert D, Bernays P (1934) *Grundlagen der Mathematik*, vol 1. Berlin (Reprinted Ann Arbor, 1944)
- Horn A (1962) The separation theorem of intuitionist propositional calculus. *J Symb Log* 27:391–399
- Kripke S (1965) Semantics analysis of intuitionistic logic. In: Crossley JN, Dummett MNE (eds) *Proceedings of the eight logic colloquium, 1963*, North Holland Publishing Co., Oxford, pp 92–130
- Łoś J (1951) An algebraic proof of completeness for the two-valued propositional calculus. *Colloq Math* 2:236–240
- Łoś J (1954) On the categoricity in power of elementary deductive systems and some related problems. *Colloq Math* 3:58–62
- Łoś J (1957) Remarks on Henkin's paper: Boolean representation through propositional calculus. *Fundam Math* 44:82–83
- Lukasiewicz J (1920) O logice trójwartościowej. *Ruch Filozoficzny* 5:169–170
- Lukasiewicz J (1929) *Elementy Logiki Matematycznej*. Warszawa (Reprinted Warszawa 1958)
- Markov AA (1950) Konstruktivnaja logika. *Usp Mat Nauk* 5:187–188
- McKinsey JCC (1941) A solution of the decision problem for the Lewis systems S.2 and S.f with an application to topology. *J Symb Log* 6:117–188
- McKinsey JCC, Tarski A (1948) Some theorems about the sentential calculi of Lewis and Heyting. *J Symb Log* 13:1–15
- Michalski RS, Larson JB (1978) Selection of most representative training examples and incremental generation of VL1 hypothesis: the underlying methodology and the description of programs ESEL and AQ11. Report No. 867, Department of Computer Science, University of Illinois, Urbana
- Moisil GC (1935) Recherches sur l'algebre de la logique. *Annales Sc de l'Univerite de Jassy* 22:1–117
- Mostowski A (1948) Proofs of non-deducibility in intuitionistic functional calculus. *J Symb Log* 13:204–207
- Nelson D (1949) Constructible falsity. *J Symb Log* 14:16–26
- Nöbeling G (1954) *Grundlagen der analitischen Topologie*, Berlin/Göttingen/Heidelberg
- Novotny N, Pawlak Z (1985) On rough equalities. *Bull Pol Acad Sci (Math)* 33:99–104
- Orłowska E (1983) Semantics of vague concepts. In: Dorn G, Weingartner P (eds) *Foundations of logic and linguistics, selected contributions to the 7th*

- International Congress of Logic, Methodology and Philosophy of Science. Plenum Press, Salzburg, pp 465–482
- Pagliani P (1994) A pure logic-algebraic analysis of rough top and rough bottom equalities. In: Rough sets, fuzzy sets and knowledge discovery, workshops in computing, Springer, pp 227–236
- Pawlak Z (1982) Rough sets. *Int J Inf Comput Sci* 11:344–356
- Pawlak Z (1985a) On learning – a rough set approach. Lecture notes in computer science, 28. Springer, Berlin, pp 197–227
- Pawlak Z (1985b) Rough sets and fuzzy sets. *Fuzzy Sets Syst* 17:99–102
- Pawlak Z (1991) Rough sets. Theory and decision library. Kluwer, Dordrecht
- Post EL (1921) Introduction to a general theory of elementary propositions. *Am J Math* 43:165–185
- Quinlan JR (1983) Learning efficient classification procedures and, their application to chess end games. In: Michalski RS, Carbonell JG, Mitchell TM (eds) Machine learning, an artificial intelligence approach. Tioga Publishing Company, Palo Alto
- Ras Z, Zemankova M (1986) On learning. A possibility approach. In: Proceedings of the 1986 'CISS, Princeton, pp 844–847
- Rasiowa H (1951) Algebraic treatment of the functional calculi of Heyting and Lewis. *Fundam Math* 38:99–126
- Rasiowa H (1974) An algebraic approach to non-classical logics. Studies in logic and the foundations of mathematics, vol 78. North-Holland Publishing Company/London – PWN, Amsterdam/Warsaw
- Rasiowa H, Cat Ho N (1992) LT-fuzzy sets. In: Zadeh LA, Kacprzyk J (eds) Fuzzy logics for management of uncertainty. Wiley, New York, pp 122–139
- Rasiowa H, Sikorski R (1950) A proof of completeness theorem of Gödel. *Fundam Math* 37:193–200
- Rasiowa H, Sikorski R (1953) Algebraic treatment of the notion of satisfiability. *Fundam Math* 40:62–95
- Rasiowa H, Sikorski R (1963) The mathematics of metamathematics. PWN, Warszawa
- Rasiowa H, Skowron A (1985) Approximation logic. In: Proceedings of mathematical methods of specification and synthesis of software systems conference. Akademie Verlag, Berlin, 31, pp 123–139
- Reichbach J (1955) On the completeness of the functional calculus of the first order. *Stud Logica* 2:245–250
- Rieger L (1951) On countable generalized σ -algebras, with, a new proof of Gödel completeness theorem. *Czechoslov Math J* 1(76):29–40
- Scott DS (1954) Prime ideal theorems for rings, lattices and, Boolean algebras. *Bull Am Math Soc* 60:390
- Sholander M (1951) Postulates for distributive lattices. *Can J Math* 3:28–30
- Sikorski R (1958) Some applications of interior mappings. *Fundam Math* 45:200–212
- Stone MH (1934) Boolean algebras and, their relation to topology. *Proc Natl Acad Sci* 20:197–202
- Stone MH (1937) Topological representation of distributive lattices and, Brouwerian logics. *Čas Mat Fys* 67:1–25
- Tarski A (1935) Grundzüge des Syntemenkalküls. Ester Teil. *Fundam Math* 25:503–526
- Tarski A (1938) Der Aussagenkalkül und die Topologie. *Fundam Math* 31:103–134
- Tarski A (1954) Prime ideal theorem for Boolean, algebras and, the axiom of choice. *Bull Am Math Soc* 60:390–391
- Tarski A, Thompson FB (1952) Some general properties of cylindric algebras. *Bull Am Math Soc* 58:65
- Thomason RH (1969) A semantical study of constructible falsity. *Z Math Logik Grundl Math* 15:247–257
- Vigneron L, Wasilewska A (1996) Rough and modal algebras. In: Proceedings of CESA'96 IMACS multi-conference: computational engineering in system applications, vol I, pp 123–130
- Vorobiev NN (1952) Konstruktivnoje iscislenie vyskasivaniij s silnym otrizaniem. *Dokl Akad Nauk SSSR* 85:456–468
- Wasilewska A (1989) Syntactic decision procedures in information systems. *Int J Man-Mach Stud* 30:273–285
- Wasilewska A (1996) On rough and LT-fuzzy sets. In: Proceedings of 1996 Asian fuzzy systems symposium – soft computing and information processing, Kenting, 11–14 Dec, pp 13–18
- Wasilewska A (1997) Topological rough algebras, Chapter 21. In: Lin TY (ed) Rough sets and database mining. Kluwer, Boston, pp 411–425
- Wasilewska A, Banerjee M (1995) Rough, sets and, topological quasi-Boolean, algebras. In: ACM CSC'95 23rd annual workshop on rough sets and database mining, conference proceedings, San Jose State University, San Jose, pp 61–67
- Wasilewska A, Vigneron L (1995) Rough, equality algebras. In: Annual joint conference on information sciences, conference proceedings, Wrightsville Beach, Oct, pp 26–30
- Wasilewska A, Vigneron L (1997) On generalized rough sets. In: Third joint conference on information sciences, intelligent systems, scientific foundations and spectrum of applications, Research Triangle Park, 2–5 Mar, pp 185–188
- Wasilewska A, Vigneron L (1998) Rough algebras and automated deduction, Chapter 14. In: Skowron A (ed) Rough set theory and its applications in various fields, in series Soft computing, Physica Verlag (division of Springer Verlag), pp 261–276
- Wiweger A (1988) On topological rough sets. *Bull Pol Acad Sci Math* 37:51–62
- Wong SKM, Ziarko W (1987) Comparison of the probabilistic approximate classification and the fuzzy set model. *Fuzzy Sets Syst* 21:357–362
- Wong SKM, Ziarko W, Ye RL (1986) On learning and evaluation of decision rules in context of rough sets. In: Proceedings of the first ACM SIGART international symposium on methodologies for intelligent systems, Knoxville, pp 308–324
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 6:338–353

Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm

Sadaaki Miyamoto
University of Tsukuba, Tsukuba, Japan

Article Outline

Introduction
Rough Approximation and Generalizations
A Set-Valued Mapping and Rough Approximations
Application to K -Means Clustering
An Illustrative Example
Conclusion
Bibliography

Introduction

Many studies have been done on rough sets (Pawlak 1982, 1991) in both theory and applications. Theoretically, a number of generalizations (e.g., Lin 1989, 1998; Slowinski and Vanderpooten 2000; Yao and Lin 1996) have been considered; on the other hand, applications are mainly focused upon information and decision tables (see, e.g., Pawlak 1991).

In this study, we consider a simple method to express upper and lower approximations of a set in both standard and a class of generalized rough approximations by using a set-valued mapping. This method is useful in both theory and applications. That is, it represents an efficient calculation method for a theoretical aspect and derives a new clustering algorithm as an application aspect.

We first briefly overview rough sets and their generalization using neighborhoods. Then a set-

valued mapping is introduced and the derivation of the upper and lower approximations as well as the rough boundary is given.

The case of an Euclidean space with the standard neighborhoods is considered and the derivation of the upper and lower approximations of a Voronoi set (Kohonen 1995; Miyamoto et al. 2008) is studied. As the K -means clustering (MacQueen 1967) uses Voronoi sets (Miyamoto et al. 2008), this derivation leads us to a class of clustering algorithm, which is similar to the rough K -means (Lingras and West 2004) and yet different from it in an important aspect.

The rest of this entry is organized as follows. Section “[Rough Approximation and Generalizations](#)” overviews rough sets and a class of generalizations. Section “[A Set-Valued Mapping and Rough Approximations](#)” introduces the formulation of rough approximations using a set-valued mapping. Section “[Application to K-Means Clustering](#)” then proposes a variation of K -means with a lower approximation of a Voronoi set. Section “[An Illustrative Example](#)” shows an illustrative example, and finally section “[Conclusion](#)” concludes the paper.

Rough Approximation and Generalizations

We begin with the classical rough approximation and then generalize it.

A universal set is denoted by X that is with a partition $\mathcal{U} = \{U_1, \dots, U_K\}$:

$$\bigcup_{i=1}^K U_i = X, \quad U_i \cap U_j = \emptyset \quad (i \neq j).$$

$U_1, \dots, U_K$ are also called equivalence classes. We mainly consider a finite number of classes but sometimes infinite cases, that is, $K = \infty$.

Given a set A of X , we define the upper approximation $R^*(A)$, the lower approximations $R_*(A)$, and the rough boundary $B(A)$:

$$R^*(A) = \bigcup_i \{U_i : A \cap U_i \neq \emptyset\}, \quad (1)$$

$$R_*(A) = \bigcup_i \{U_i : U_i \subseteq A\}. \quad (2)$$

$$B(A) = R^*(A) - R_*(A) \quad (3)$$

It is straightforward to assume $K = \infty$ here: infinite number of U_i .

Well-known properties of these approximations are as follows: (Pawlak 1982)

$$R^*(\overline{A}) = \overline{R_*(A)}, \quad (4)$$

$$R^*(A \cup B) = R^*(A) \cup R^*(B) \quad (5)$$

$$R^*(A \cap B) \subseteq R^*(A) \cap R^*(B) \quad (6)$$

$$R_*(A \cup B) \supseteq R_*(A) \cup R_*(B) \quad (7)$$

$$R_*(A \cap B) = R_*(A) \cap R_*(B) \quad (8)$$

The proofs are easy and omitted (see also the next section).

Partition $\mathcal{U}$ is equivalent to an equivalence relation, say R . R is a binary relation (i.e., $R(x, y) = 1$ or $R(x, y) = 0$) defined on $X \times X$ which satisfies the following three properties:

(i) [Reflexive property]

$$R(x, x) = 1 \text{ for all } x \in X.$$

(ii) [Symmetric property]

$$R(x, y) = R(y, x) \text{ for all } x, y \in X.$$

(iii) [Transitive property]

$$\text{If } R(x, y) = 1 \text{ and } R(y, z) = 1, \text{ then } R(x, z) = 1 \text{ for all } x, y, z \in X.$$

Given a partition $\mathcal{U} = \{U_1, \dots, U_K\}$, the corresponding equivalence relation R is defined:

— $R(x, y) = 1$ iff x and y are in a same class, that is, $x, y \in U_i$ for some U_i .

— $R(x, y) = 0$ iff x and y are in different classes, that is, $x \in U_i$ and $y \in U_j$ for $i \neq j$.

Conversely, given an equivalence relation R , an equivalence class is defined by

$$U_x = \{y \in X : R(x, y) = 1\}. \quad (9)$$

For $x, y \in X$, either

$$U_x = U_y \quad (10)$$

Or

$$U_x \cap U_y = \emptyset \quad (11)$$

holds.

To prove this, suppose $z \in U_x \cap U_y$. Then $R(x, z) = 1$ and $R(z, y) = 1$, which implies $R(x, y) = 1$ by the transitive property. Thus, $y \in U_x$. For any $w \in U_y$, we have $w \in U_x$ from the same property, and hence, $U_y \subseteq U_x$. In the same way, we have $U_x \subseteq U_y$. Hence, $U_x = U_y$.

We thus see that to give a partition and an equivalence relation is the same.

A Class of Generalizations

Let us relax the condition (i)–(iii) of the equivalence relation. The condition (iii) is more complicated than others, and to remove this is natural. Thus, we assume

(i) R is reflexive ($R(x, x) = 1, \forall x \in X$) and

(ii) R is symmetric ($R(x, y) = R(y, x), \forall x, y \in X$)

in this subsection.

We consider how R is related to a set like (9). We use the symbol N_x instead of U_x for this generalization:

$$N_x = \{y \in X : R(x, y) = 1\}, \quad (12)$$

when R satisfies (i) and (ii).

It is immediate to see that (i) implies

$$x \in N_x$$

and (ii) implies

$$y \in N_x \Leftrightarrow x \in N_y.$$

However, Eqs. (10) and (11) no longer hold. We call N_x with the above properties a *neighborhood* of x

Note 1 The neighborhood herein is different from the concept of neighborhood in the standard mathematical sense, as we consider only one neighborhood N_x to an object x , while an infinite family of sets including x is considered in the standard concept of neighborhood.

The upper and lower approximation using N_x is as follows:

$$R^*(A) = \bigcup_{x \in X} \{N_x : A \cap N_x \neq \emptyset\}, \quad (13)$$

$$R_*(A) = \bigcup_{x \in X} \{x : N_x \subseteq A\}. \quad (14)$$

$$B(A) = R^*(A) - R_*(A) \quad (15)$$

Note that $R_*(A) \neq \bigcup_{x \in X} \{N_x : N_x \subseteq A\}$ in general. If $N_x = U_x$ forms a partition (i.e., equivalence classes), then we have

$$\begin{aligned} R_*(A) &= \bigcup_{x \in X} \{x : U_x \subseteq A\} \\ &= \bigcup_{x \in X} \{U_x : U_x \subseteq A\}. \end{aligned} \quad (16)$$

Proposition 1 The relations (4, 5, 6, 7, and 8) also holds for (13) and (14) defined by N_x .

Proof Suppose $\overline{A} \cap N_x \neq \emptyset$. Then $N_x \subseteq A$ does not hold. Hence $x \in R_*(A)$. It is easy to see that the converse is also true. Hence, $R^*(\overline{A}) = \overline{R_*(A)}$.

The proof of $R^*(A \cup B) = R^*(A) \cup R^*(B)$ and $R^*(A \cap B) \subseteq R^*(A) \cap R^*(B)$ will be shown in the next section, since it is a special case of the properties of mappings in general. The last two relations are easily derived by using the above three relations:

$$\begin{aligned} R_*(A \cap B) &= \overline{R^*(\overline{A \cap B})} = \overline{R^*(\overline{A} \cap \overline{B})} \\ &= \overline{R^*(\overline{A}) \cup R^*(\overline{B})} = \overline{R^*(\overline{A})} \cap \overline{R^*(\overline{B})} \\ &= R_*(A) \cap R_*(B). \end{aligned}$$

$$\begin{aligned} R_*(A \cup B) &= \overline{R^*(\overline{A \cup B})} = \overline{R^*(\overline{A} \cap \overline{B})} \\ &\supseteq \overline{R^*(\overline{A}) \cap R^*(\overline{B})} = \overline{R^*(\overline{A})} \cap \overline{R^*(\overline{B})} \\ &= R_*(A) \cup R_*(B). \end{aligned}$$

□

Note that this proposition also holds for the classical rough approximations, since the condition for R is less restrictive than the equivalence relation for the classical case.

Note 2 Our formulation is somewhat loose in handling infinity from the view-point of rigorous mathematics, that is, “countable infinity” and “uncountable infinity” are not distinguished. Specifically, U_x and N_x can be used for uncountable cases, while U_i not. We can note that uncountable cases can be handled like the countable infinity by appropriately changing notations. Details are omitted.

A Set-Valued Mapping and Rough Approximations

Assume again that a partition $\mathcal{U}$ is given on X . In other words, R is an equivalence relation.

Let us introduce a natural projection $p : X \rightarrow X/\mathcal{U}$

$$p(x) = [x]_{\mathcal{U}}, \quad (17)$$

In other words, $p(x) = U_i \in \mathcal{U}$ such that $x \in U_i$. Note that $p(x)$ is not a subset of X . We hence need another mapping $q(x)$:

$$q(x) = p^{-1}(p(x)), \quad (18)$$

In other words, $q(x) = U_i \subset X$ such that $x \in U_i$. Thus, $q(x)$ is a set-valued mapping: $q : X \rightarrow 2^X$.

Note 3 $p^{-1}(\cdot)$ is the inverse image and not the inverse mapping. Specifically, for a given $V \in X/\mathcal{U}$,

$$p^{-1}(V) = \{x \in X : p(x) \in V\}.$$

Using $q(x)$, the upper approximation and lower approximation are given by:

$$R^*(A) = \bigcup_{x \in A} q(x) = q(A), \quad (19)$$

$$B(A) = \bigcup_{x \in R^*(A)-A} q(x) = q(q(A) - A), \quad (20)$$

$$\begin{aligned} R_*(A) &= R^*(A) - B(A) \\ &= q(A) - q(q(A) - A). \end{aligned} \quad (21)$$

It is immediate to see that the approximation using Eqs. (19) and (21) coincide with the standard definition of Eqs. (1) and (2), respectively.

The above mapping has the significance of efficient calculation of the lower approximation, assuming the upper approximation is straightforward using $q(A)$. Unlike the standard theory of rough sets, the rough boundary is secondly calculated and the lower approximation is lastly calculated.

Suppose the direct calculation of $R_*(A)$ seems inefficient when compared with the upper approximation. Moreover, if X is huge, then to take the complement $\bar{A}$ is difficult. In such a case, the above calculation using $q(\cdot)$ is useful.

The Case of Neighborhood

We proceed to the case of N_x using a set-valued mapping, in other words, relation R is reflexive and symmetric. The set-valued mapping is straightforward:

$$q(x) = N_x \quad (22)$$

Thus, the upper approximation, $R^*(A)$ the lower approximation $R_*(A)$, and the rough boundary $B(A)$ are defined by $q(x) = N_x$. Specifically,

$$R^*(A) = \bigcup_{x \in A} N_x, \quad (23)$$

$$B'(A) = \bigcup_{x \in R^*(A)-A} N_x, \quad (24)$$

$$\begin{aligned} R_*(A) &= R^*(A) - B'(A) \\ &= \bigcup_{x \in A} N_x - \bigcup_{x \in R^*(A)-A} N_x, \end{aligned} \quad (25)$$

$$B(A) = R^*(A) - R_*(A). \quad (26)$$

Note that $B'(A)$ is not the rough boundary, since $B'(A) \subseteq R^*(A)$ does not hold.

It now is straightforward to prove $R^*(A \cup B) = R^*(A) \cup R^*(B)$ and $R^*(A \cap B) \subseteq R^*(A) \cap R^*(B)$, since they are, respectively, special forms of $q(A \cup B) = q(A) \cup q(B)$ and $q(A \cap B) \subseteq q(A) \cap q(B)$, and the last two are well-known relations of set mapping.

In the following examples, first example handles an Euclidean space where the Euclidean metric is given and a neighborhood is generated from the metric, while the second example shows the opposite direction: a relation is given and a metric is derived from the relation.

Example 1 A typical example is that $X \subseteq \mathbf{R}^p$ where $\mathbf{R}^p$ is a p -dimensional Euclidean space and

$$N_x = \{y \in X : \|y - x\| \leq d\}, \quad (27)$$

where $\|y - x\|$ is the Euclidean distance between y and x and d is a positive parameter. The correspondence between R and N_x by Eq. (27) is given by:

$$xRy \Leftrightarrow \|x - y\| \leq d. \quad (28)$$

Example 2 Suppose X is finite for simplicity. R is given on X which is reflexive and symmetric. Accordingly we can define N_x as above. X can be visualized using an undirected graph: An edge $\{x, y\}$ exists in the graph if and only if $R(x, y) = 1$. Then X is regarded as a metric space (X, D) where metric D is defined to be the minimum number of edges between x and y except that $D(x, x) = 0$ for all $x \in X$. When x and y are not connected by a number of edges:

$D(x, y) = +\infty$. It is trivial to see $D(x, y) = D(y, x)$ from the symmetry of R . The triangular inequality is also easy to prove, since $D(x, y) + D(y, z)$ cannot be less than $D(x, z)$: the minimum number of edges connecting x and z . If we assume that all x and y in X are connected, $D(x, y) < +\infty$.

Example 3 Suppose that R is transitive in Example 2. In this case, the generalized rough approximations are reduced to the classical rough approximations using a partition. The graph is now a set of complete subgraphs, where x and y are either connected by an edge or not connected at all. Thus, $D(x, y) = 1$ or $D(x, y) = < +\infty$, respectively.

Example 4 In real applications, we rarely have a transitive relation, while reflexive and symmetric relations are commonly found. For example, SNS users are forming symmetric relation, and reflexive property is regarded trivial. Hence, assume an undirected graph $G = (X, E)$ in which X and E are, respectively, sets of vertices and edges; assume also G is not a complete graph. A relation R is the same as edges $E : R(x, y) = 1$ iff $\{x, y\} \in E$. A problem related to transitive property is to find another graph $G' = (X, E')$ such that $E \subseteq E'$ and the corresponding relation R' is transitive. Moreover, R' should be minimal in the sense that if there is another graph $G'' = (X, E'')$ with the same property, $E' \subseteq E''$. Such a relation R' is called a *transitive closure* of R . It is easy to find G' : first assume $E' = E$ and repeat for all pair $\{x, y\} \in E'$, $\{y, z\} \in E'$ and $\{x, z\} \notin E'$, add edges $\{x, z\}$ and $\{z, x\}$ to E' until no edge can be added. (This means that “a friend of a friend is also a friend.”) As the result of this algorithm, connected components of G are made into complete subgraphs. (A connected component of G is a subgraph whose vertices are reachable by tracing edges and cannot be reachable to other vertex of the same graph.)

Note 4 The transitive closure can be applied to metric spaces including Euclidean spaces: it is closely related to the single linkage clustering

(Everitt 1993; Miyamoto 1990). The details are omitted, since the topic is different from the discussion herein.

Soft Boundary in an Euclidean Space

Let us assume that a simple set of a semi space in R^p is given:

$$A = \{y \in X : \langle y, a \rangle + \beta \leq 0\}, \quad (29)$$

where $a \in R^p$ is a given vector and β is a scalar. We can assume $\|a\| = 1$ without loss of generality. The boundary of A is $\langle y, a \rangle + \beta = 0$.

Assume that the neighborhood N_x is defined by (27). Then it is immediate to see that the rough approximation is given as follows.

$$R^*(A) = \{y \in X : \langle y, a \rangle + \beta \leq d\}, \quad (30)$$

$$B^l(A) = \{y \in X : -d \leq \langle y, a \rangle + \beta \leq 2d\}, \quad (31)$$

$$R_*(A) = \{y \in X : \langle y, a \rangle + \beta \leq -d\} \quad (32)$$

$$B(A) = \{y \in X : -d \leq \langle y, a \rangle + \beta \leq d\}. \quad (33)$$

Thus $B(A)$ is the region between the two hyperplanes, which can be called as a soft boundary between A and its complement $\bar{A}$.

Note 5 Rigorous mathematical discussion should distinguish open and closed sets, while their distinction is not essential in some applications. Hence, the closures of sets are taken when necessary in the next application of clustering (like $Cl(R_*(A))$), where closed sets are easier to handle than open ones.

Application to K-Means Clustering

Clustering using the concepts of rough sets has been studied by several researchers (Lingras and West 2004; Kinoshita et al. 2014; Ubukata et al. 2016). They all consider variations of K -means (MacQueen 1967; Bezdek 1981). A drawback in these studies is that the upper and lower

approximations are defined in a heuristic way and the region of clusters does not have a simple rule of classification, while K -means have Voronoi regions (Kohonen 1995; Miyamoto et al. 2008), and fuzzy c -means have simple classification functions (Miyamoto et al. 2008).

Let us assume that $X = \{x_1, \dots, x_N\}$ is a set of objects for clustering, and x_k ($k = 1, \dots, N$) is a point in $\mathbf{R}^p$. Clusters are denoted by G_i ($i = 1, \dots, c$) which are subsets of X . Each object x_k belongs to a unique G_i . Hence, we assume:

$$\bigcup_{i=1}^c G_i = X; G_i \cap G_j = \emptyset \ (i \neq j).$$

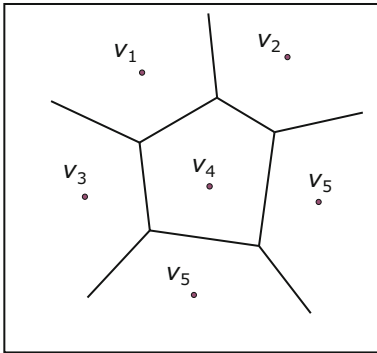
Given cluster centers $V = (v_1, \dots, v_c)$, a Voronoi region $V(v_i)$ is given by

$$V(v_i) = \{x \in \mathbf{R}^p : \|x - v_i\| \leq \|x - v_j\|, \forall v_j \neq v_i\}.$$

Note that $V(v_i)$ is a subset of $\mathbf{R}^p$, unlike G_i which is a subset of X . Figure 1 is an example of Voronoi regions in a plane.

K-Means Algorithm

The best-known clustering algorithm is the K -means (MacQueen 1967). Although there are variations, the basic algorithm is a simple iteration of the calculation of centroids of clusters and the allocation of objects to their nearest centers (see, e.g., Bezdek 1981; Miyamoto et al. 2008).



Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm, Fig. 1 An example of Vornoi regions with centers v_i ($1 \leq i \leq 6$) in a plane

Assume that v_i is the cluster center of G_i . The K -means algorithm is as follows.

Basic K-means Algorithm.

Step 0. Determine initial cluster centers v_i ($1 \leq i \leq c$).

Step 1. For $1 \leq k \leq N$, allocate x_k to its nearest center:

$$x_k \in G_i \Leftrightarrow i = \arg \min_{1 \leq j \leq c} \|x_k - v_j\|^2. \quad (34)$$

Step 2. Calculate new cluster centers:

$$v_i = \frac{1}{|G_i|} \sum_{x_k \in G_i} x_k, 1 \leq i \leq c, \quad (35)$$

where $|G_i|$ is the number of objects in G_i .

Step 3. Test if the clusters are convergent. If not, go to **Step 1**.

End Basic K-means.

It is well known that the method of K -means is reformulated as an alternative minimization algorithm of an objective function (Bezdek 1981):

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^N u_{ki} \|x_k - v_i\|^2, \quad (36)$$

where $V = (v_1, \dots, v_c)$ and $U = (u_{ki})$. V is a matrix whose columns are cluster centers and U is an $N \times c$ matrix with the meaning:

$$u_{ki} = \begin{cases} 1 & (x_k \in G_i), \\ 0 & (x_k \notin G_i), \end{cases} \quad (37)$$

with the constraint $\sum_{i=1}^c u_{ki} = 1$ for all $1 \leq k \leq N$.

That is, x_k belongs to one and only one cluster.

The alternative minimization algorithm is the following:

Algorithm AM.

AMO. Give initial cluster centers $\bar{V}$.

AMI. Solve $\min_U J(U, \bar{V})$ and let the optimal solution be $\bar{U}$.

AM2. Solve $\min_V J(\bar{U}, V)$ and let the optimal solution be $\bar{V}$.

AM3. Test if the solution $(\bar{U}, \bar{V})$ is convergent. If not, go to **AM1**.

End AM.

It is easy to see that the optimal solution of **AM1** and **AM2** are:

$$u_{ki} = 1 \Leftrightarrow i = \arg \min_{1 \leq j \leq c} \|x_k - v_j\|^2, \quad (38)$$

$$u_i = \frac{\sum_{k=1}^N u_{ki} x_k}{\sum_{k=1}^N u_{ki}}. \quad (39)$$

They are obviously the same as (34) and (42), respectively. The objective function (36) can be rewritten as

$$J(U, V) = \sum_{i=1}^c \sum_{x_k \in G_i} \|x_k - v_i\|^2. \quad (40)$$

Note that the alternative minimization does not mean the strict optimization of $J(U, V)$ with respect to (U, V) . In other words, the solution of **AM** after convergence is not the global minimum of $J(U, V)$ in general. Hence, different initial values will lead to different solutions. In practice, a number of different initial values are tried and the solution of the minimum value of $J(U, V)$ among them is adopted as the best solution. Note also that the objective function value monotonically decreases in the iteration of **AM1-AM3**. If the solution changes, the objective function value will decrease. This monotone decreasing property can be used as a convergence criterion as shown below.

Convergence test in the K-means. Convergence test in the above **K-means** algorithm (including **AM**) can be done in different ways. First, difference in the memberships can be tested: if all the memberships are the same as the last ones, the algorithm terminates. Second, the convergence of the centers can be tested. If all the

locations of the centers do not change, the algorithm terminates. In addition, the objective function value can be used as a convergence criterion. If the change in the objective function value is zero or negligible, the algorithm stops.

Allocation Rule

The Eqs. (34) and (38) imply a generic allocation rule:

$$x \in \text{cluster } i \Leftrightarrow i = \arg \min_{1 \leq j \leq c} \|x - v_j\|^2 \quad (41)$$

defined for every $x \in \mathbf{R}^p$ regardless whether $x \in X$ or $x \notin X$. The symbol cluster i , which implies a region where a new $x \in \mathbf{R}^p$ should be allocated to that cluster, is distinguished from G_i . It is easy to see cluster $i = V(v_i)$, a Voronoi region.

Note 6 It is true that the above allocation rule does not work on the boundary of Voronoi regions. Moreover, if a point in X is on the boundary, its allocation to a cluster is not uniquely determined. We assume that such a point is exceptional and does not influence the whole discussion and omit detailed discussion. If necessary, random allocation will be sufficient.

Rough Approximations

Rough K -means by Lingras (Lingras and West 2004) use two different weights w_1 and w_2 with $w_1 + w_2 = 1$ and $w_1 > w_2 > 0$ and calculates cluster center v_i as follows:

$$v_i = \frac{w_1}{|L(G_i)|} \sum_{x_k \in L(G_i)} x_k + \frac{w_2}{|B(G_i)|} \sum_{x_l \in B(G_i)} x_l, \quad (42)$$

where $|L(G_i)|$ and $|B(G_i)|$ are, respectively, a lower approximation and a rough boundary of cluster G_i which are different from $R^*(G_i)$ and $R_*(G_i)$ in the following.

There are two points for more consideration in this method (Lingras and West 2004). First, the upper and lower approximations are not clearly related to a rough approximation and seem *ad hoc*.

Second, the method is not formulated as an alternative optimization, unlike the K -means and fuzzy c -means (see, e.g., Bezdek 1981).

The above consideration of a clustering algorithm leads us to another algorithm in which the calculation of cluster center is similar to Eq. (42), the case of rough K -means, while we use the above approximation using the neighborhood (27).

Note that the boundary of K -means clusters is that of a Voronoi region which consist of hyperplanes:

$$\|x - v_j\|^2 - \|x - v_i\|^2 = 0 \quad (43)$$

which is reduced to

$$\begin{aligned} L(x; v_i, v_j) &= 2\langle x, v_i + v_j \rangle + \|v_j\|^2 \\ &\quad - \|v_i\|^2 \\ &= 0. \end{aligned} \quad (44)$$

We thus define

$$L_{ij} = \{x \in \mathbf{R}^p : L(x; v_i, v_j) = 0\}. \quad (45)$$

The distance between a point y and the hyperplane L_{ij} by (44) is given by

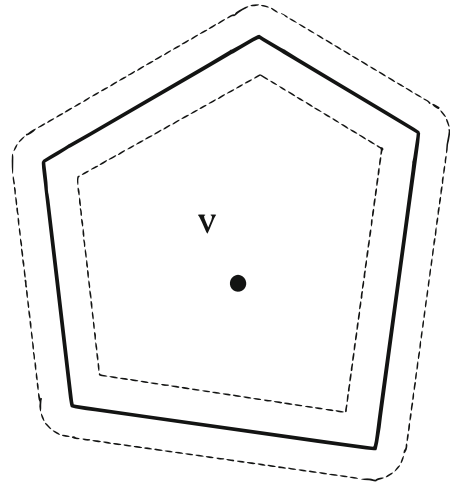
$$\begin{aligned} \text{Dist}(y, L_{ij}) &= \frac{|L(y; v_i, v_j)|}{2\|v_i + v_j\|} \\ &= \frac{|2\langle y, v_i + v_j \rangle + \|v_j\|^2 - \|v_i\|^2|}{2\|v_i + v_j\|}. \end{aligned} \quad (46)$$

Note that $G_i = V(v_i) \cap X$. Hence, we define the regions of $\mathbf{R}^p$ $U(v_i)$ and $L(v_i)$, respectively, corresponding to $R^*(G_i)$ and $R_*(G_i)$. Thus,

$$R^*(G_i) = U(v_i) \cap X, \quad (47)$$

$$R_*(G_i) = L(v_i) \cap X. \quad (48)$$

The upper and lower approximations for G_i in two-dimensional case are in three regions shown in Fig. 2.



Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm, Fig. 2 The two regions $U(v_i)$ and $L(v_i)$ shown by dotted lines with a Voronoi region $V(v_i)$ shown by solid lines, where $L(v_i) \subset V(v_i) \subset U(v_i)$

The section of the set-valued mapping showed that the upper approximation has a simpler representation than the lower approximation. In contrast, the lower approximation $R_*(G_i)$ has a simpler calculation formula than the upper approximation, that is, the following conditions should be checked for $x \in \mathbf{R}^p$.

$$x \in V(v_i) \Leftrightarrow i = \arg \min_{1 \leq j \leq c} \|x - v_j\|, \quad (49)$$

$$\text{Dist}(x, L_{ij}) \geq d, \forall 1 \leq j \leq c, j \neq i. \quad (50)$$

If Eqs. (49) and (50) are satisfied, then $x \in X$ is in $R_*(G_i)$.

In contrast, to judge if x is in $R^*(G_i)$ is not simple. Note the boundary of $U(v_i)$ in Fig. 2 has linear parts and curved parts. Moreover, an upper approximation of a cluster does not seem to be useful in clustering, since clusters basically mean a family of disjoint subsets (Everitt 1993). In K -means, disjoint subsets are easily derived from the Voronoi sets.

A lower approximation on the other hand can be used to distinguish a point in a cluster that is near to a boundary or not. The latter is also useful

to make an algorithm to be more robust by changing the weight of contribution of a point.

Therefore, we consider a cluster G_i and its lower approximation $R_*(G_i)$, and not $R^*(G_i)$. In the following algorithm, we find a cluster G_i :

$$G_i = V_i(v_i) \cap X \quad (51)$$

when a cluster center is given.

The lower approximation is determined:

$$R_*(G_i) = \{x \in G_i : \text{Dist}(y, L_{ij}) \geq d\}. \quad (52)$$

and the next cluster center is calculated by:

$$v_i = \frac{w_1}{|R_*(G_i)|} \sum_{x_k \in R_*(G_i)} x_k + \frac{w_2}{|G_i - R_*(G_i)|} \sum_{x_l \in G_i - R_*(G_i)} x_l. \quad (53)$$

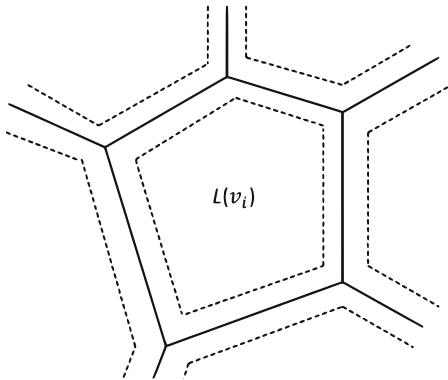
using the idea of the rough K -means. An example of Voronoi regions with the lower approximations (and without an upper approximation) is shown in Fig. 3.

The proposed algorithm (Miyamoto et al. 2018) is thus as follows.

ARKM Algorithm (Another Rough K -Means)

Step 1: Set initial values of $v_i (i = 1, \dots, c)$

Step 2: Find G_i by (51) and $R_*(G_i)$ by (52).



Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm, Fig. 3 Voronoi regions with the lower approximations $L(v_i)$

Step 3: Calculate new $v_i (i = 1, \dots, c)$ by Eq. (53).

Step 4: If cluster centers are convergent, stop; else go to **Step 2**.

End ARKM.

Note that the more specific calculation of G_i is easy:

$$x_k \in G_i \Leftrightarrow \|x_k - v_i\| \leq \|x_k - v_j\|, \forall j \neq i. \quad (54)$$

Moreover,

$$x_k \in R_*(G_i) \Leftrightarrow x_k \in G_i \text{ and } \text{Dist}(x_k, L_{ij}) \geq d \quad (55)$$

Optimization Problem

This algorithm actually is equivalent to an alternative optimization of an objective function. Let us consider the following:

$$J(\mathcal{G}, \mathcal{G}', V) = w_2 J_1(\mathcal{G}, V) + (w_1 - w_2) J_2(\mathcal{G}', V), \quad (56)$$

$$J_1(\mathcal{G}, V) = \sum_{x_k \in G_i} \|x_k - v_i\|^2, \quad (57)$$

$$J_2(\mathcal{G}', V) = \sum_{x_k \in G_i, N_{x_k} \subseteq V(v_i)} \|x_k - v_i\|^2, \quad (58)$$

where

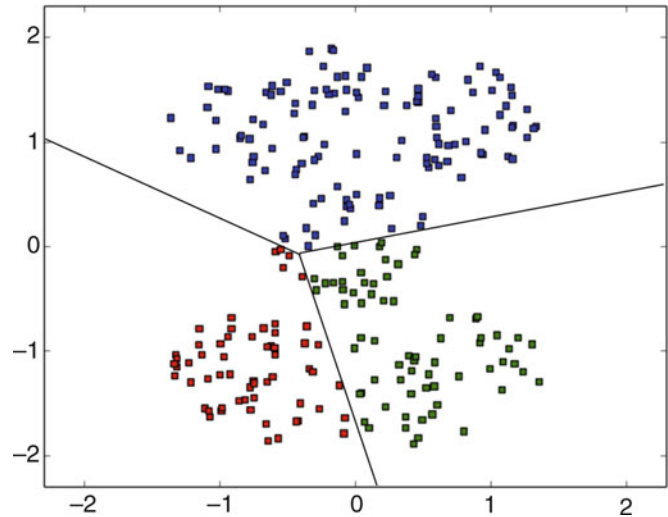
$$\mathcal{G} = \{G_1, \dots, G_c\}, \mathcal{G}' = \{G'_1, \dots, G'_c\}$$

are partitions of X .

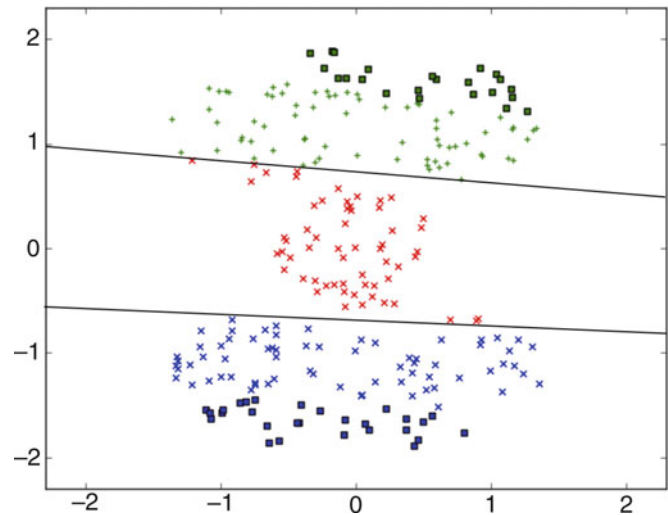
Proposition 2 The solutions of **ARKM** actually minimizes the objective function $J(\mathcal{G}, \mathcal{G}', V)$ in the sense of alternative optimization.

Proof Consider the alternative optimization of $J(\mathcal{G}, \mathcal{G}', V)$: then it is easy to see that we can take $\mathcal{G} = \mathcal{G}'$ and u_i is given by Eq. (53), since objects in $V(v_i)$ should be allocated to the same G_i even when J_2 is concerned. $\square$

Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm, Fig. 4 Result of K -means with $K = 3$ derived from “synthetic cassini”



Set-Valued Mapping for Generalized Rough Approximation and Application to Clustering Algorithm, Fig. 5 Result from ARKM algorithm with $K = 3$, $w_2/w_1 = 0.6$. Two symbols $+$ and $\times$ are those objects in $G_i - R_*(G_i)$; other symbols using squares indicate objects in $R_*(G_i)$. The middle cluster does not have the lower approximation



Note 7 The ARKM algorithm and the alternative optimization are shown in (Miyamoto et al. 2018).

An Illustrative Example

Let us show an illustrative example whereby how ARKM algorithm work. The data is called “synthetic cassini” (<https://clusteval.sdu.dk/1/datasets>). It consists of two prolonged banana-like clusters and a round cluster.

A result from K -means is shown in Fig. 4 where we do not find a prolonged cluster; the smallest cluster was separated into three. In

contrast, the result in Fig. 5 shows better clusters. Two symbols $+$ and $\times$ are those objects in $G_i - R_*(G_i)$; other symbols using squares indicate objects in $R_*(G_i)$. The middle cluster does not have the lower approximation $R_*(G_i)$.

The example here is only for the purpose of illustration, but more results are shown in (Miyamoto et al. 2018).

Conclusion

We have proposed a new formulation of rough approximation using set-valued mappings. The

classical rough approximations using equivalence relation and the case of neighborhoods were considered. Networks and Euclidean spaces are contrasted.

As an application, a method of K -means clustering using rough approximations are proposed, where only the lower approximation, and not the upper approximation, was used.

There are further research possibilities in both theory and applications. Fuzzy neighborhoods and fuzzy equivalence relations can be handled using the present framework. Other metrics such as the L_1 space should be considered with application to clustering should be considered. Moreover, fuzzy clustering can be studied.

Bibliography

- Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum, New York
- Everitt BS (1993) Cluster analysis, 3rd edn. Arnold, London
- Kinoshita N, Endo Y, Miyamoto S (2014) On some models of objective-based rough clustering. In: Proceedings of 2014 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT), Warsaw
- Kohonen T (1995) Self-organizing maps. Springer, Berlin
- Lin TY (1989) Neighborhood systems and approximation in relational databases and knowledge bases. In: Proceedings of the fourth international symposium on methodologies of intelligent systems, Poster Session, pp 75–86
- Lin TY (1998) Granular computing on binary relations I: data mining and neighborhood systems. In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery 1. Physica-Verlag, Heidelberg, pp 107–121
- Lingras P, West C (2004) Interval set clustering of web users with rough k-means. *J Intell Inf Syst* 23:5–16
- MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press, Berkeley, California, pp 281–297
- Miyamoto S (1990) Fuzzy sets in information retrieval and cluster analysis. Springer, Heidelberg
- Miyamoto S, Ichihashi H, Honda K (2008) Algorithms for fuzzy clustering. Springer, Berlin
- Miyamoto S, Choi JM, Endo Y, Huynh VN (2018) Optimal clustering with twofold memberships. In: Torra V et al. (eds) Proceedings of MDAI2018, LNAI. Springer Nature, Switzerland AG
- Pawlak Z (1982) Rough sets. *Int J Comput Inf Sci* 11:341–356
- Pawlak Z (1991) Rough sets-theoretical aspects of reasoning about data. Kluwer, Dordrecht
- Slowinski R, Vanderpooten D (2000) A generalized definition of rough approximations based on similarity. *IEEE Trans Knowl Data Eng* 12(2):331–336
- Ubukata S, Notsu A, Honda K (2016) The rough set k-means clustering. *Proc SCIS-ISIS 2016*:189–193
- Yao YY, Lin TY (1996) Generalization of rough sets using modal logics. *Intell Autom Soft Comput* 2(2):103–120

Fuzzy System Representations of Stochastic Systems

Hong-Xing Li

School of Control Science and Engineering,
Dalian University of Technology, Dalian,
P. R. China

Article Outline

Glossary
Definition of the Subject
Introduction
Background of Stochastic Systems
Fuzzy Reasoning Meaning of Continuous Stochastic Systems
Fuzzy Reasoning Representations of Double-Inputs and Single-Output Continuous Stochastic Systems
Fuzzy Reasoning Representations of Discrete Stochastic Systems
Reducibility in the Transformations Between Fuzzy Systems and Stochastic Systems
The Uncertainty Systems with One-Dimensional Random Variable and Their Representations
Unification on Uncertainty Systems
Conclusions
Bibliography

Glossary

Fuzzy systems The systems modeled by means of a group of fuzzy inference rules that are formed by using fuzzy sets.

Stochastic systems The systems modeled by means of a kind of probability density functions that describe relations between the input variables and the output variables of the systems.

Uncertain systems The systems with uncertain variables that may be fuzzy variables or random variables or having the both.

Unification of Fuzzy systems and Stochastic systems A fuzzy system can be turned into a stochastic system or a stochastic system can be turned into a fuzzy system.

Definition of the Subject

An uncertain system may be a fuzzy system or a stochastic system. To consider the relationship between a fuzzy system and a stochastic system is surely an important and interesting problem. Fuzzy system representations of stochastic systems mean that a stochastic system can be turned into a fuzzy system. Of course, stochastic system representations of fuzzy systems mean that a fuzzy system can be turned into a stochastic system.

Introduction

Uncertainty systems have played more and more important role in many areas such as control theory, system engineering, artificial intelligence, behavior science, social science, etc. As we all know, uncertainty of systems are usually with respect to randomness or fuzziness. So if people focus on randomness of an uncertainty system, they must use probability theory or stochastic process to describe the system, which uncertainty system is called a stochastic system; if people pay attention to fuzziness of an uncertainty system, they often treat the system by using fuzzy set theory, which uncertainty system is called a fuzzy system. Clearly, it is interesting to communicate the relationship between probability theory and fuzzy set theory with respect to uncertainty systems. This entry will research the relationship between probability theory and fuzzy set theory

when they are used to deal with uncertainty systems.

Background of Stochastic Systems

We consider Fig. 1 that shows a single-input single-output open-loop system S . We have known that if this system S is a deterministic system then one may use the conventional method to make a mathematical model of the system and find a solution $y(x)$ of the model by analytic or numerical methods. In this way, this system shall be regarded as having been mastered basically. Then the system S may be simply understood by a function, denoted by a mapping as follows, where the set X is the input universe and the set Y is the output universe:

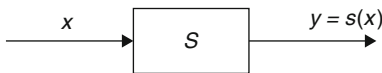
$$s : X \rightarrow Y, \quad x \mapsto y \triangleq s(x).$$

The system is formally denoted by $s = S(X, Y)$.

However, for an uncertain system, we cannot use the conventional method to get a “crisp” or “accurate” function s . Having to reduce the request, we try to obtain an approximate function $\underline{s}$ as follows:

$$\underline{s} : X \rightarrow Y, \quad x \mapsto y \triangleq \underline{s}(x), \quad (1)$$

such that $\underline{s}(x)$ approximates $s(x)$ as close as possible. With stochastic viewpoint to understand above problem there is such meaning: randomly take a point x in X and put into the input channel of system S , and after enter S there exists an output point $y(x)$ in the output channel of system S to correspond the input point x . However, what point in Y should be taken as $y(x)$ is unknown beforehand. This means that for system S there are two



Fuzzy System Representations of Stochastic Systems, Fig. 1 Single-input single-output open-loop system

random variables ξ and η that are defined, respectively, in probability spaces $(X, \mathcal{B}_1, P_1)$ and $(Y, \mathcal{B}_2, P_2)$ where $\mathcal{B}_1$ and $\mathcal{B}_2$ are σ -fields on X and Y respectively, and P_1 and P_2 are probability measures on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively.

For convenience, we always assume that X and Y are measurable sets on real number space $\mathbb{R}$. Evidently η and ξ depend on each other, that is, there is a Borel measurable function g such that $\eta = g(\xi)$, which ought to coincide with $y = s(x)$, i.e., $\eta = s(\xi)$. We want to determine a Borel measurable function $\bar{s}(\xi)$ so that η and $\bar{s}(\xi)$ are closed up to the best, and then $\bar{s}$ may be regarded as an approximation of s , where the existence of $E(\eta^2)$ and $E[(s(\xi))^2]$ are assumed. The “closeness” herein needs a criterion, and the most commonly used one is “least squares method.” Then we have to demand that

$$E[(\eta - \underline{s}(\xi))^2] = \inf_{\varphi} \left\{ E[(\eta - \varphi(\xi))^2] \right\},$$

where φ varies in a kind of space of Borel measurable functions. We have known that conditional mathematical expectation should meet the demand; so we can put $\underline{s}(\xi) \triangleq E(\eta|\xi)$. This shows that random variable $\bar{s}(\xi)$ is the optimal approximation in mean square to random variable η . But, it is only of formal meaning as we have not got any probability information about random vector (ξ, η) . Now we start to do the work.

Taking $\Omega \triangleq X \times Y$, $\mathcal{F} \triangleq \mathcal{B}_1 \times \mathcal{B}_2$, and $P \triangleq P_1 \times P_2$, where $\mathcal{F}$ is Borel σ -field generated by Cartesian product of Borel σ -fields $\mathcal{B}_1$ and $\mathcal{B}_2$, and P is product probability measure, then we obtain joint probability space $(\Omega, \mathcal{F}, P)$. With the same notations, redefine ξ and η as random variables on Ω :

$$\begin{aligned} \xi : \Omega &\rightarrow \mathbb{R}, & (x, y) &\mapsto \xi(x, y) \triangleq \xi(x), \\ \eta : \Omega &\rightarrow \mathbb{R}, & (x, y) &\mapsto \eta(x, y) \triangleq \eta(x). \end{aligned}$$

Thus (ξ, η) turns into a two-dimensional random vector on joint probability space $(\Omega, \mathcal{F}, P)$. For any $x \in X$, when $\omega \in \{\omega \in \Omega | \xi = x\}$, we have

$$\underline{s}(x) = E(\eta|\xi = x). \quad (2)$$

This means that $\underline{s}(x)$ becomes the conditional mathematical expectation of random variable η under condition of random variable $\xi = x$.

If we master whole probability information on (ξ, η) , especially know continuous probability density $f(x, y)$ of (ξ, η) , then (2) turns into the following equation which is easily handled:

$$\underline{s}(x) = E(\eta|\xi = x) = \frac{\int_{-\infty}^{+\infty} f(x, y)ydy}{\int_{-\infty}^{+\infty} f(x, y)dy}, \quad (3)$$

where the demands need that, for any $x \in X$,

$$\begin{aligned} \int_{-\infty}^{+\infty} |y| f(x, y)dy &< +\infty, \\ 0 &< \int_{-\infty}^{+\infty} f(x, y)dy < +\infty. \end{aligned}$$

Clearly in actual computing, (3) should be the following:

$$\underline{s}(x) = \frac{\int_y f(x, y)ydy}{\int_y f(x, y)dy}. \quad (4)$$

For an uncertainty system S , if we can know the continuous probability density $f(x, y)$ on the system, $\underline{s}$ defined by (4) is called a **continuous stochastic approximation system** of S , or a **continuous stochastic system** of S . But in our mind, we should know that the **continuous stochastic system** $\underline{s}$ is with regard to an uncertainty system S . So after time when we say a continuous stochastic system $\underline{s}$, it always has the above meaning. Besides for convenience, a continuous stochastic system $\underline{s}$ is also denoted as the following:

$$\underline{s} = S(X, Y, f(x, y)) \quad (5)$$

Be careful that (4) and (5) use the same symbol $\underline{s}$, which means that $\underline{s}$ has two meanings: it not only abstractly expresses a continuous stochastic approximation system, but also indicates correspondence relationship between X and Y .

Fuzzy Reasoning Meaning of Continuous Stochastic Systems

For convenience, we need to introduce some concepts. Given a universe X , $\mathcal{A} = \{A_i | 1 \leq i \leq n\}$ is a family of normal fuzzy sets on X , i.e.,

$$(\forall i \in \{1, 2, \dots, n\})(\exists x_i \in X)(\mu_{A_i}(x_i) = 1),$$

where x_i is called peak point of A_i ; of course, the peak point is not unique. $\mathcal{A}$ is called a fuzzy partition of X , if it meets the condition:

$$(\forall x \in X) \left(\sum_{i=1}^n \mu_{A_i}(x) = 1 \right). \quad (6)$$

It is not difficult to verify that such fuzzy sets have Kronecker property:

$$\mu_{A_i}(x_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Besides, for proving the following main theorem, we have to give two lemmas. These two lemmas are with respect to integral with parameter.

Lemma 1 Let $f(x, y)$ be a binary continuous function on $X \times Y$, where $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are finite real number intervals. For the integral with parameter $I(x) = \int_{a_2}^{b_2} f(x, y)dy$, we have such a result: for arbitrarily given $\varepsilon > 0$, there is always a common $\delta > 0$ without depending on parameter x such that, for any partition:

$$a_2 = y_0 < y_1 < \dots < y_n = b_2,$$

as long as $\lambda = \max \{\Delta y_i | i = 1, 2, \dots, n\} < \delta$, then the Riemann sum $\sum_{i=1}^n f(x, \xi_i) \Delta y_i$ of $I(x)$ uniformly holds that

$$\left| I(x) - \sum_{i=1}^n f(x, \xi_i) \Delta y_i \right| < \varepsilon$$

for all $x \in X$, where $\Delta y_i = y_i - y_{i-1}$ ($i = 1, 2, \dots, n$) and ξ_i takes its value in $[y_{i-1}, y_i]$ arbitrarily.

Proof For arbitrarily given $\varepsilon > 0$, let

$$\delta_k = \frac{1}{k}, k = 1, 2, \dots.$$

We can prove that there must exist a k such that $\delta = \delta_k$ meets the conclusion of the lemma. If it is not, then for every k , exist $x_k \in X$ and a partition:

$$a_2 = y_0^{(k)} < y_1^{(k)} < \dots < y_{n_k}^{(k)} = b_2$$

of Y and a kind of taking value way of $\xi_i^{(k)} \in [y_{i-1}^{(k)}, y_i^{(k)}]$, although

$$\lambda_k = \max \left\{ \Delta y_i^{(k)} \mid i = 1, 2, \dots, n_k \right\} < \delta_k,$$

we have $\left| I(x_k) - \sum_{i=1}^{n_k} f(x_k, \xi_i^{(k)}) \Delta y_i^{(k)} \right| \geq \varepsilon$. As $\{x_k\}$ is a bounded sequence, it has a convergent subsequence as being $\{x_{k_j}\}$ such that $x_{k_j} \xrightarrow{j \rightarrow \infty} x_*$. Noticing the fact:

$$\delta_{k_j} \xrightarrow{j \rightarrow \infty} 0,$$

so we have that

$$\begin{aligned} 0 < \varepsilon &\leq \lim_{j \rightarrow \infty} \left| I(x_{k_j}) - \sum_{i=1}^{n_{k_j}} f(x_{k_j}, \xi_i^{(k_j)}) \Delta y_i^{(k_j)} \right| \\ &= \left| I(x_*) - \int_{a_2}^{b_2} f(x_*, y) dy \right| = 0. \end{aligned}$$

This is a clear contradiction. **Q.E.D.**

Lemma 2 Let $f(x, y)$ be a binary continuous function on $X \times Y$, where $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are finite real number intervals. For the integral with parameter $I(x) = \int_{a_2}^{b_2} f(x, y) dy$, if $(\forall x \in X)(I(x) > 0)$, then there exists a $\delta > 0$, such that, for any partition:

$$a_2 = y_0 < y_1 < \dots < y_n = b_2$$

of Y and any kind of taking value way of ξ_i in the interval $[y_{i-1}, y_i]$, the Riemann sum $\sum_{i=1}^n f(x, \xi_i) \Delta y_i$ of the integral $I(x)$ must hold that if the following inequality is true

$$\lambda \triangleq \max \{ \Delta y_i \mid i = 1, 2, \dots, n \} < \delta,$$

Then we have the result:

$$(x \in X) \left(\sum_{i=1}^n f(x, \xi_i) \Delta y_i > 0 \right)$$

Proof It is a well-known fact that

$$I(x) = \int_{a_2}^{b_2} f(x, y) dy \in C[a_1, b_1].$$

Thus, there exists the least point of $x_0 \in X$ for $I(x)$ in the universe X such that

$$(\forall x \in X)(I(x) \geq I(x_0)).$$

Take $\varepsilon = I(x_0)$. From Lemma 1 we know that there exists a $\delta > 0$ such that for any partition of Y

$$a_2 = y_0 < y_1 < \dots < y_n = b_2$$

and any kind of taking value way of ξ_i in $[y_{i-1}, y_i]$ the Riemann sum $\sum_{i=1}^n f(x, \xi_i) \Delta y_i$ of $I(x)$ must meet the condition:

$$\lambda < \delta \Rightarrow (x \in X) \left(\left| I(x) - \sum_{i=1}^n f(x, \xi_i) \Delta y_i \right| < \varepsilon \right).$$

Then we have the following expression:

$$\sum_{i=1}^n f(x, \xi_i) \Delta y_i > I(x) - \varepsilon \geq I(x_0) - \varepsilon = 0.$$

This is uniformly true for all $x \in X$. **Q.E.D.**

Based on Lemma 2 and by using the way in the proof of Lemma 1, we can easily prove the following lemma.

Lemma 3 Let $f(x, y)$ be a binary continuous function on $X \times Y$, where $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are finite real number intervals, and it is with the condition:

$$(\forall x \in X) \left(\int_{a_2}^{b_2} f(x, y) dy > 0 \right).$$

For arbitrarily given $\varepsilon > 0$, there is always a common $\delta > 0$ without dependent of parameter x such that, for any partition:

$$a_2 = y_0 < y_1 < \cdots < y_n = b_2,$$

as long as $\lambda = \max \{ \Delta y_i | i = 1, 2, \dots, n \} < \delta$, then

$$\left| \frac{\int_{a_2}^{b_2} f(x, y) y dy}{\int_{a_2}^{b_2} f(x, y) dy} - \frac{\sum_{i=1}^n f(x, \xi_i) \xi_i \Delta y_i}{\sum_{i=1}^n f(x, \xi_i) \Delta y_i} \right| < \varepsilon$$

is uniformly true for all $x \in X$, where

$$\Delta y_i = y_i - y_{i-1}, i = 1, 2, \dots, n$$

and ξ_i takes its value in $[y_{i-1}, y_i]$ arbitrarily. **Q.E.D.**

Theorem 1 Given a continuous stochastic system

$$\underline{g} = S(X, Y, f(x, y)),$$

where $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are two finite real number intervals. If

$$(\forall x \in X) \left(\int_{a_2}^{b_2} f(x, y) dy > 0 \right),$$

then there exists a group of fuzzy inference rules:

$$\text{If } x \text{ is } A_i \text{ then } y \text{ is } B_i, i = 1, 2, \dots, n, \quad (7)$$

where $A_i \in \mathcal{F}(X)$ and $B_i \in \mathcal{F}(Y)$ such that the fuzzy system $\bar{s}$ constructed by the group of fuzzy inference rules can approximate the continuous stochastic system $\underline{g}$ to arbitrarily given precision.

Proof Noticing (4), we make a partition of interval Y as the following:

$$a_2 = y_0 < y_1 < \cdots < y_n = b_2,$$

and write $\gamma \triangleq (y_1, y_2, \dots, y_n)$ and

$$\begin{aligned} \Delta y_i &= y_i - y_{i-1}, i = 1, 2, \dots, n, \\ \lambda &= \max \{ \Delta y_i | i = 1, 2, \dots, n \} \end{aligned}$$

Then we make the Riemann sum $\sum_{i=1}^n f(x, y_i) y_i \Delta y_i$

and $\sum_{i=1}^n f(x, y_i) \Delta y_i$ of $f(x, y)y$ and $f(x, y)$ on Y , respectively, with respect to the partition and the node group γ . From the condition of the theorem and Lemma 2, we have that, $\delta_1 > 0$, if $\lambda < \delta_1$, then

$$(x \in X) \left(\sum_{i=1}^n f(x, y_i) \Delta y_i > 0 \right).$$

So we have the following expression:

$$\begin{aligned} \underline{g}(x) &= E(\eta | \xi = x) = \frac{\int_{a_2}^{b_2} f(x, y) y dy}{\int_{a_2}^{b_2} f(x, y) dy} \\ &\approx \frac{\sum_{i=1}^n f(x, y_i) y_i \Delta y_i}{\sum_{i=1}^n f(x, y_i) \Delta y_i} = \sum_{i=1}^n \left(\frac{f(x, y_i) \Delta y_i}{\sum_{j=1}^n f(x, y_j) \Delta y_j} \right) y_i \\ &= \sum_{i=1}^n \mu_{A_i^*}(x) y_i, \end{aligned}$$

where we have made a definition:

$$\mu_{A_i^*}(x) \triangleq \frac{f(x, y_i) \Delta y_i}{\sum_{j=1}^n f(x, y_j) \Delta y_j}$$

and clearly $A_i^* \in \mathcal{F}(X)$. Now for any given approximation precision $\varepsilon > 0$, because $f(x, y)$ is continuous, by using lemma 3, $\exists \delta_2 > 0$ and $\delta_2 < \delta_1$, when $\lambda < \delta_2$, for all $x \in X$ the following holds uniformly:

$$\left| \underline{g}(x) - \sum_{i=1}^n \mu_{A_i^*}(x) y_i \right| < \frac{\varepsilon}{2}. \quad (8)$$

Based on the nodes y_j ($j = 0, 1, \dots, n$), we make $n + 1$ fuzzy sets B_j ($j = 0, 1, \dots, n$) on Y with demand of B_j being a fuzzy partition on Y and continuous on Y , which is regarded as a kind of fuzziness of these real numbers y_j ($j = 0, 1, \dots, n$); for example, B_j can be taken as “triangular waves” membership functions:

$$\begin{aligned} \mu_{B_0}(y) &= \begin{cases} (y - y_1)/(y_0 - y_1), & y_0 \leq y \leq y_1, \\ 0, & \text{otherwise;} \end{cases} \\ \mu_{B_j}(y) &= \begin{cases} (y - y_{j-1})/(y_j - y_{j-1}), & y_{j-1} \leq y \leq y_j, \\ (y - y_{j+1})/(y_j - y_{j+1}), & y_j < y \leq y_{j+1}, \\ 0, & \text{otherwise,} \end{cases} \\ j &= 1, 2, \dots, n-1; \\ \mu_{B_n}(y) &= \begin{cases} (y - y_{n-1})/(y_n - y_{n-1}), & y_{n-1} \leq y \leq y_n, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

And make $A_i = (i = 1, \dots, n)$ as the following:

$$\mu_{A_i}(x) \triangleq f(x, y_i) / \sum_{j=1}^n f(x, y_j). \quad (9)$$

Let $\mathcal{A} \triangleq \{A_i | 1 \leq i \leq n\}$ and $\mathcal{B} \triangleq \{B_i | 1 \leq i \leq n\}$. Regarding $\mathcal{A}$ and $\mathcal{B}$ as linguistic variables that take their values in themselves, we form a group of fuzzy inference rules as the same as (7):

If x is A_i then y is B_i , $i = 1, 2, \dots, n$.

Now we construct a fuzzy system by means of CRI method as follows.

First, from i -th fuzzy inference rule of (7), a fuzzy relation $R_i \triangleq A_i \times B_i$ on $X \times Y$ is formed, where its membership function is

$$(\forall (x, y) \in X \times Y) (\mu_{R_i}(x, y) = \mu_{A_i}(x) \wedge \mu_{B_i}(y)).$$

Since the n fuzzy inference rules should be combined by “or,” a whole fuzzy inference relation $R = \bigcup_{i=1}^n R_i$ is obtained, i.e., for any $(x, y) \in X \times Y$,

$$\mu_R(x, y) = \bigvee_{i=1}^n \mu_{R_i}(x, y) = \bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y)).$$

For any $A \in \mathcal{F}(X)$, a fuzzy inference result $B \in \mathcal{F}(Y)$ should be got by R , which is equivalent to a fuzzy transformation from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$ being introduced by R , denoted by “ $\circ$ ”, i.e.,

$$\circ : \mathcal{F}(X) \rightarrow \mathcal{F}(Y), \quad A \rightarrow B = \circ(A) \triangleq A \circ R,$$

where its membership function is as follows

$$\mu_B(y) = \bigvee_{x \in X} (\mu_A(x) \wedge \mu_R(x, y)), \quad y \in Y. \quad (10)$$

For arbitrarily given an input $x' \in X$, for we may use (10), x' should be turned into a fuzzy set $A' \in \mathcal{F}(X)$ as $\mu_{A'}(x) \triangleq \chi_{\{x'\}}(x)$, where χ_D is the characteristic function of a set D . Substituting A' into (10), we obtain a result of reasoning $B' \in \mathcal{F}(Y)$ as follows

$$\mu_{B'}(y) = \bigvee_{i=1}^n (\mu_{A_i}(x') \wedge \mu_{B_i}(y)), \quad y \in Y \quad (11)$$

Since B' is a fuzzy set, we have to obtain exact quantity $y' \in Y$ by a defuzzification technique. From (11) we know that $\mu_{B'}(y)$ is piecewise continuous on Y . So we have that

$$\begin{aligned} \int_{a_2}^{b_2} |y| \mu_{B'}(y) dy &< +\infty, \\ \int_{a_2}^{b_2} \mu_{B'}(y) dy &< +\infty. \end{aligned}$$

Now we can prove that $\int_{a_2}^{b_2} \mu_{B'}(y) dy > 0$. In fact, if it is not true, i.e., $\int_{a_2}^{b_2} \mu_{B'}(y) dy = 0$, then since $\mu_{B'}(y)$ is piecewise continuous on Y and nonnegative, we have that $\mu_{B'}(y) = 0$, a.e. Y . Because $\sum_{i=1}^n f(x', y_i) > 0$, there exists $i_0 \in \{1, 2, \dots, n\}$ such that $\mu_{A_{i_0}}(x') > 0$. Noticing (11), we know that $\mu_{B'}(y) = 0$, a.e. Y , which is contrary with the definition of all B_j being continuous normal fuzzy sets. Thus, we can put

$$y' = \int_{a_2}^{b_2} y \mu_{B'}(y) dy / \int_{a_2}^{b_2} \mu_{B'}(y) dy.$$

Then we get the correspondence point y' in Y of x' . By the arbitrariness of x' , x' can be replaced by a general point x in X , and y' is replaced by $\bar{s}(x)$. And we obtain a function $\bar{s}: X \rightarrow Y$ as follows

$$\bar{s}(x) \triangleq \frac{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y)) \right] y dy}{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y)) \right] dy}. \quad (12)$$

$\bar{s}$ is a fuzzy approximation system with respect to the uncertainty system S , denoted by the following symbol:

$$\bar{s} \triangleq S(X, Y, \mathcal{A}, \mathcal{B}).$$

Besides, it is not difficult to understand that

$$(\forall x \in X) \left(\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y)) \right] dy > 0 \right).$$

By Lemma 2, we have that, $\exists \delta_3 > 0$, such that when $\lambda > \delta_3$, for any $x \in X$,

$$\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y_j)) \right] \Delta y_j > 0.$$

Noticing the Riemann sun of (12) and B_i being with Kronecker property, we have

$$\begin{aligned} \bar{s}(x) &\approx \frac{\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y_j)) \right] y_j \Delta y_j}{\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{A_i}(x) \wedge \mu_{B_i}(y_j)) \right] \Delta y_j} \\ &= \frac{\sum_{j=1}^n \mu_{A_j}(x) y_j \Delta y_j}{\sum_{j=1}^n \mu_{A_j}(x) \Delta y_j} = \frac{\sum_{j=1}^n f(x, y_j) y_j \Delta y_j}{\sum_{j=1}^n f(x, y_j) \Delta y_j} \\ &= \sum_{i=1}^n \left(\frac{f(x, y_i) \Delta y_i}{\sum_{j=1}^n f(x, y_j) \Delta y_j} \right) y_i = \sum_{i=1}^n \mu_{A_i^*}(x) y_i, \end{aligned} \quad (13)$$

where we have put that

$$\mu_{A_i^*}(x) \triangleq \frac{f(x, y_i) \Delta y_i}{\sum_{j=1}^n f(x, y_j) \Delta y_j}.$$

From (13) and Lemma 3, we know that $\exists \delta_4 > 0$ and $\delta_4 > \delta_3$, when $\lambda < \delta_4$, for all $x \in X$, the following holds uniformly:

$$\left| \bar{s}(x) - \sum_{i=1}^n \mu_{A_i^*}(x) y_i \right| < \frac{\varepsilon}{2}.$$

At last, take $\delta = \min \{\delta_2, \delta_4\}$. When $\lambda < \delta$, for all $x \in X$, the following holds uniformly:

$$\begin{aligned} &|\bar{s}(x) - \underline{s}(x)| \\ &= \left| \bar{s}(x) - \sum_{i=1}^n \mu_{A_i^*}(x) y_i + \sum_{i=1}^n \mu_{A_i^*}(x) y_i - \underline{s}(x) \right| \\ &\leq \left| \bar{s}(x) - \sum_{i=1}^n \mu_{A_i^*}(x) y_i \right| + \left| \sum_{i=1}^n \mu_{A_i^*}(x) y_i - \underline{s}(x) \right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This means that fuzzy system $\bar{s}$ can approximate continuous stochastic system $\underline{s}$ to arbitrarily given ε . **Q.E.D.**

Example 1 Given a continuous stochastic system, where $X = Y = (-\infty, \infty)$ and $f(x, y)$ is a binary normal probability density $N(a_1, a_2, \sigma_1^2, \sigma_2^2, r)$, i.e.,

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right]}$$

If we take these numbers as follows:

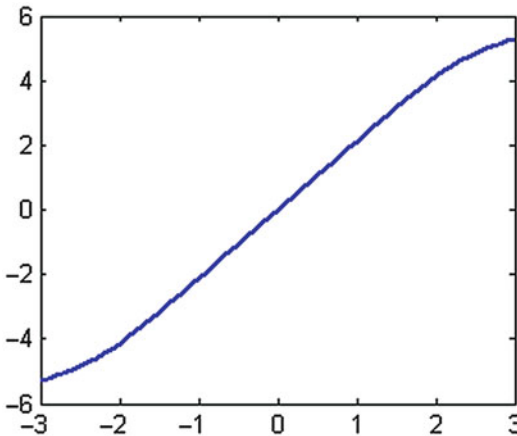
$$r^2 = 0.5, a_1 = 0, a_2 = 0, \sigma_1 = 0.5, \sigma_2 = 1.5,$$

then the above equation is that

$$f(x, y) = \frac{2\sqrt{2}}{3\pi} e^{-\left(4x^2 - \frac{4\sqrt{2}}{3}xy + \frac{4}{9}y^2\right)}.$$

Based on 3σ principle, X and Y can be approximately taken as finite intervals. For example, X is taken as the interval $[-3, 3]$ by $6\sigma_1$, and Y is taken as $[-6, 6]$ by $4\sigma_2$. By theorem 1, we have the following equation and its image can refer to Fig. 2.

$$\begin{aligned} \underline{g}(x) &= E(\eta|\xi = x) \\ &= \frac{\int_{-6}^6 y \exp\left[-\left(4x^2 - \frac{4\sqrt{2}}{3}xy + \frac{4}{9}y^2\right)\right] dy}{\int_{-6}^6 \exp\left[-\left(4x^2 - \frac{4\sqrt{2}}{3}xy + \frac{4}{9}y^2\right)\right] dy}. \end{aligned}$$



Fuzzy System Representations of Stochastic Systems, Fig. 2 Input cure of the system $\underline{g}(x)$

Now we consider the fuzzy approximation system $\bar{s}$. First make a partition of Y :

$$y_0 = -6, y_1 = -5, y_2 = -4, \dots, y_{11} = 5, y_{12} = 6;$$

then make fuzzy sets B_j ($j = 0, 1, \dots, 12$) as in Fig. 3. $A_i(x)$ ($i = 0, 1, \dots, 12$) are defined as the following equation (see (9)), their images refer to Fig. 4.

$$\mu_{A_i}(x) = \frac{f(x, y_i)}{\sum_{j=1}^{12} f(x, y_j)} = \frac{e^{-\left(4x^2 - \frac{4\sqrt{2}}{3}xy_i + \frac{4}{9}y_i^2\right)}}{\sum_{j=1}^{12} e^{-\left(4x^2 - \frac{4\sqrt{2}}{3}xy_j + \frac{4}{9}y_j^2\right)}}$$

We now calculate that

$$\bar{s}(x) = \frac{\int_{-6}^6 \left[\bigvee_{i=1}^{12} (\mu_{A_i}(x) \wedge \mu_{B_i}(y)) \right] y dy}{\int_{-6}^6 \left[\bigvee_{i=1}^{12} (\mu_{A_i}(x) \wedge \mu_{B_i}(y)) \right] dy},$$

which image refers to Fig. 5. And the images of $\bar{s}(x)$ and $\underline{g}(x)$ refer to Fig. 6.

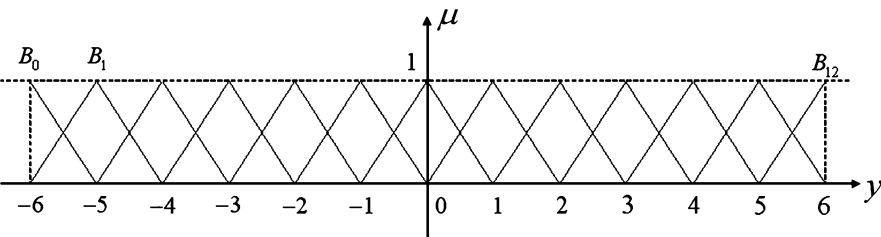
If the partition nodes $\{y_j\}$ are increased by double number, i.e.,

$$y_0 = -6, y_1 = -5.5, y_2 = -5, \dots, y_{24} = 6,$$

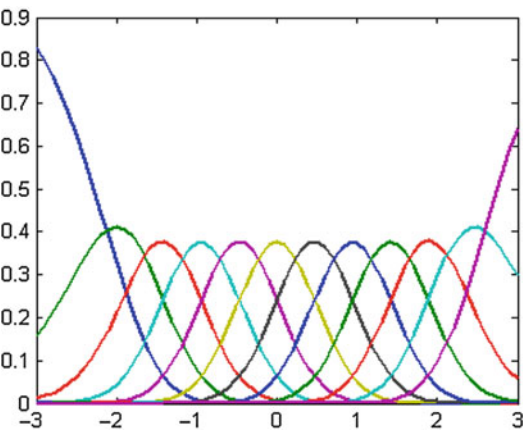
then the approximation precision increases better, where the comparison cures of $\bar{s}(x)$ and $\underline{g}(x)$ refer to Fig. 7. And if the partition nodes $\{y_j\}$ are increased, then the cures of $\bar{s}(x)$ and $\underline{g}(x)$ are basically coincident from Fig. 8.

Fuzzy Reasoning Representations of Double-Inputs and Single-Output Continuous Stochastic Systems

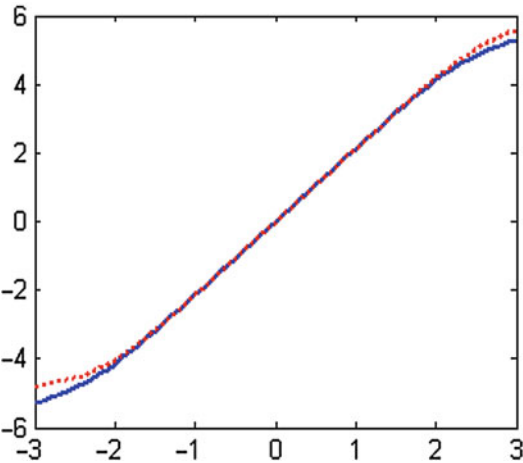
Figure 9 shows a double-input single-output open-loop system S . The input variables x and y take values in the input universes X and Y , respectively, and the output variable z takes values in the output universe Z . If this system S is a deterministic system then one may use the



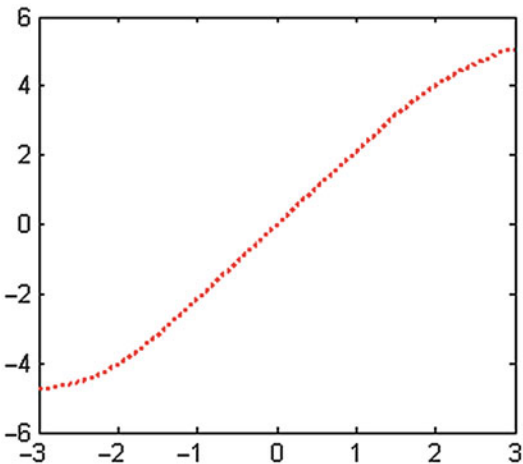
Fuzzy System Representations of Stochastic Systems, Fig. 3 Membership functions of B_j



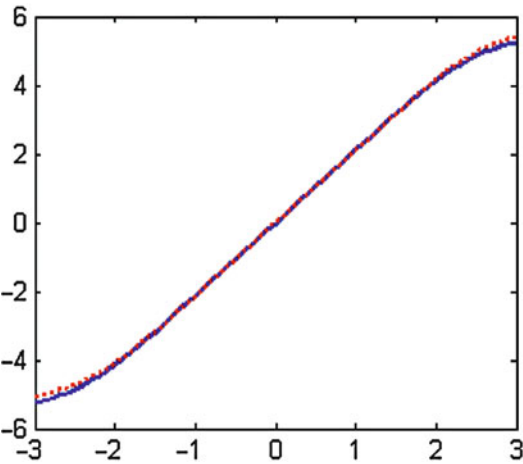
Fuzzy System Representations of Stochastic Systems, Fig. 4 Membership functions of A_i



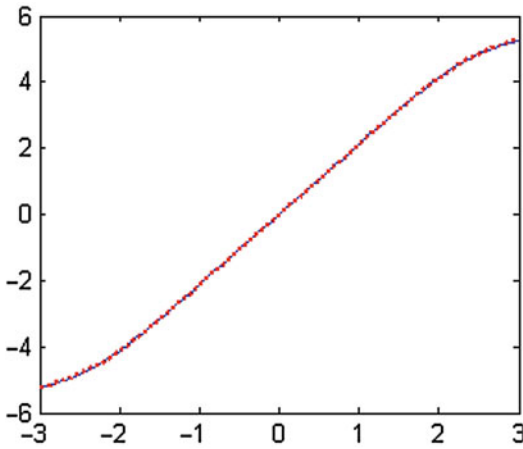
Fuzzy System Representations of Stochastic Systems, Fig. 6 The comparison of $\bar{s}(x)$ and $\underline{s}(x)$ when $\lambda = 1$, where “...” indicates $\bar{s}(x)$ and “—” represents $\underline{s}(x)$



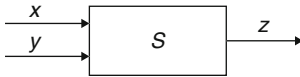
Fuzzy System Representations of Stochastic Systems, Fig. 5 Output cure of fuzzy system $\bar{s}$



Fuzzy System Representations of Stochastic Systems, Fig. 7 The comparison of $\bar{s}(x)$ and $\underline{s}(x)$ when $\lambda = 0.5$, where “...” indicates $\bar{s}(x)$ and “—” represents $\underline{s}(x)$



Fuzzy System Representations of Stochastic Systems, Fig. 8 The comparison of $\bar{s}(x)$ and $s(x)$ when $\lambda = 0.1$, where “...” indicates $\bar{s}(x)$ and “—” represents $s(x)$



Fuzzy System Representations of Stochastic Systems, Fig. 9 A double-input single-output open-loop system

conventional method to make a mathematical model of the system S (for example, one can use the mechanism modeling approach to establish a partial differential equation model) and find a solution $z(x, y)$ of the model by analytic or numerical methods. In this way, this system shall be regarded as having been mastered basically.

Then the system S may be simply understood by a binary function, also denoted by S , i.e.,

$$s : X \times Y \rightarrow Z, (x, y) \mapsto z \triangleq s(x, y). \quad (14)$$

When S is an uncertainty system, although it is hard to get a precise function as (14), we may often obtain an approximate function as follows

$$\underline{s} : X \times Y \rightarrow Z, (x, y) \mapsto z \triangleq \underline{s}(x, y), \quad (15)$$

such that $\underline{s}(x, y)$ and $s(x, y)$ are very close. Just as stated in [Background of Stochastic Systems](#), we still consider to realize above approximation thought by using conditional mathematical expectation.

Let X , Y , and Z be measurable sets on real number space $\mathbb{R}$, and ξ , η and ζ be random variables defined on probability spaces $(X, \mathcal{B}_1, P_1)$, $(Y, \mathcal{B}_2, P_2)$, and $(Z, \mathcal{B}_3, P_3)$ respectively, where $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ are Borel σ -fields on X , Y , and Z , respectively, and P_1 , P_2 , and P_3 are probability measure on $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$, respectively. Taking $\Omega \triangleq X \times Y \times Z$ and

$$\mathcal{F} \triangleq \mathcal{B}_1 \times \mathcal{B}_2 \times \mathcal{B}_3, P \triangleq P_1 \times P_2 \times P_3$$

we obtain joint probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is the Borel σ -field generated by Cartesian product of Borel σ -fields $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$, and P is the product probability measure. With same notations, redefine ξ , η , and ζ as random variables on Ω :

$$\begin{aligned} \xi : \Omega &\rightarrow \mathbb{R}, (u, v, w) \mapsto \xi(u, v, w) \triangleq \xi(u), \\ \eta : \Omega &\rightarrow \mathbb{R}, (u, v, w) \mapsto \eta(u, v, w) \triangleq \eta(v), \\ \zeta : \Omega &\rightarrow \mathbb{R}, (u, v, w) \mapsto \zeta(u, v, w) \triangleq \zeta(w). \end{aligned}$$

Then a three-dimensional random vector (ξ, η, ζ) defined on $(\Omega, \mathcal{F}, P)$ is got. For any a point $(x, y) \in X \times Y$, when $\omega \in \{\omega \in \Omega \mid \xi = x, \eta = y\}$, let

$$\underline{s}(x, y) \triangleq E(\zeta \mid \xi = x, \eta = y) \quad (16)$$

As stated in [Background of Stochastic systems](#), (16) is the optimal approximation in mean square to $s(x, y)$.

Suppose that we have known the continuous probability density $f(x, y, z)$ of the random vector (ξ, η, ζ) . Then the conditional mathematical expectation (16) can be written as follows:

$$\underline{s}(x, y) = \frac{\int_Z f(x, y, z) z dz}{\int_Z f(x, y, z) dz}, \quad (17)$$

where there is a demand: for any $(x, y) \in X \times Y$,

$$\begin{aligned} \int_Z |z| f(x, y, z) dz &< +\infty, \\ 0 &< \int_Z f(x, y, z) dz < +\infty. \end{aligned}$$

When the continuous probability density $f(x, y, z)$ of the uncertainty system is known, $\underline{g}$ defined by (17) is called a double-input single-output continuous approximation stochastic system or simply a double-input single-output continuous approximation stochastic system, denoted by

$$\underline{g} = S(X \times Y, Z, f(x, y, z)). \quad (18)$$

For proving the following main theorem, we still need two lemmas.

Lemma 4 Let $f(x, y, z)$ be a ternary continuous function on $X \times Y \times Z$, where $X = [a_1, b_1]$, $Y = [a_2, b_2]$, and $Z = [a_3, b_3]$ are finite real number intervals. For the integral with parameter

$$I(x, y) = \int_{a_3}^{b_3} f(x, y, z) dz,$$

we have such a result: for arbitrarily given $\varepsilon > 0$, there is always a common $\delta > 0$ without dependent of parameter (x, y) such that, for any partition of Z :

$$a_3 = z_0 < z_1 < \cdots < z_n = b_3,$$

as long as $\lambda = \max \{ \Delta z_i | i = 1, 2, \dots, n \} < \delta$, then the Riemann sum $\sum_{i=1}^n f(x, y, \xi_i) \Delta z_i$ of $I(x, y)$ must meet the condition that

$$\left| I(x, y) - \sum_{i=1}^n f(x, y, \xi_i) \Delta z_i \right| < \varepsilon$$

is uniformly true for all $(x, y) \in X \times Y$, where

$$\Delta z_i = z_i - z_{i-1} (i = 1, 2, \dots, n)$$

and ξ_i takes its value in $[z_{i-1}, z_i]$ arbitrarily.

Proof For arbitrarily given $\varepsilon > 0$, let

$$\delta_k = \frac{1}{k}, k = 1, 2, \dots$$

We can prove that there must exist one k such that $\delta = \delta_k$ meets the conclusion of the lemma. If it is not, then for every k , exist $(x_k, y_k) \in X \times Y$ and a partition of Z as follows

$$a_3 = z_0^{(k)} < z_1^{(k)} < \cdots < z_{n_k}^{(k)} = b_3$$

and a kind of taking value way of $\xi_i^{(k)}$ in $[z_{i-1}^{(k)}, z_i^{(k)}]$, although

$$\lambda_k = \max \{ \Delta z_i^{(k)} | i = 1, 2, \dots, n_k \} < \delta_k,$$

we have the following inequality:

$$\left| I(x_k, y_k) - \sum_{i=1}^{n_k} f(x_k, y_k, \xi_i^{(k)}) \Delta z_i^{(k)} \right| \geq \varepsilon. \quad (19)$$

Notice that $\{x_k\}$ in binary sequence $\{(x_k, y_k)\}$ are all bounded sequences and so it has a convergent subsequence $\{x_{k_j}\}$ such that $x_{k_j} \rightarrow x_* \in X (j \rightarrow \infty)$. In the same way, for $\{y_{k_j}\}$ there is a convergent subsequence $\{y_{k_{j_p}}\}$ such that $y_{k_{j_p}} \xrightarrow{p \rightarrow \infty} y_* \in Y$.

And noticing $\delta_{k_{j_p}} \xrightarrow{p \rightarrow \infty} 0$, so we have

$$0 < \varepsilon \leq \lim_{p \rightarrow \infty}$$

$$\begin{aligned} & \left| I(x_{k_{j_p}}, y_{k_{j_p}}) - \sum_{i=1}^{n_{k_{j_p}}} f(x_{k_{j_p}}, y_{k_{j_p}}, \xi_i^{(k_{j_p})}) \Delta z_i^{(k_{j_p})} \right| \\ &= \left| I(x_*, y_*) - \int_{a_3}^{b_3} f(x_*, y_*, z) dz \right| = 0. \end{aligned}$$

This is a clear contradiction. **Q.E.D.**

Lemma 5 Let $f(x, y, z)$ be a ternary continuous function in the set $X \times Y \times Z$, where $X = [a_1, b_1]$, $Y = [a_2, b_2]$ and $Z = [a_3, b_3]$ are finite real number intervals. For the integral with parameter as follows

$$I(x, y) = \int_{a_3}^{b_3} f(x, y, z) dz,$$

if $(\forall(x, y) \in X \times Y)(I(x, y) > 0)$, then there exists a $\delta > 0$, such that for any partition of Z :

$$a_3 = z_0 < z_1 < \cdots < z_n = b_3$$

and any kind of taking value way of ξ_i in $[z_{i-1}, z_i]$, the Riemann sum $\sum_{i=1}^n f(x, y, \xi_i) \Delta z_i$ of $I(x, y)$ must meet the following condition: if the following inequality holds

$$\lambda = \max \{ \Delta z_i | i = 1, 2, \cdots, n \} < \delta,$$

then we have the result:

$$(\forall(x, y) \in X \times Y) \left(\sum_{i=1}^n f(x, y, \xi_i) \Delta z_i > 0 \right).$$

Because the way of the proof is the same as the proof of Lemma 4, it is omitted. **Q.E.D.**

Besides, we have the following lemma just like Lemma 3.

Lemma 6 Let $f(x, y, z)$ be a ternary continuous function in $X \times Y \times Z$, where $X = [a_1, b_1]$, $Y = [a_2, b_2]$ and $Z = [a_3, b_3]$ are finite real number intervals, and meet the condition:

$$(\forall(x, y) \in X \times Y) \left(\int_{a_3}^{b_3} f(x, y, z) dz > 0 \right).$$

For arbitrarily given $\varepsilon > 0$, there is always a common $\delta > 0$ without dependent of parameter (x, y) such that, for any partition of Z as follows

$$a_3 = z_0 < z_1 < \cdots < z_n = b_3,$$

as long as $\lambda = \max \{ \Delta z_i | i = 1, 2, \cdots, n \} < \delta$, then

$$\left| \frac{\int_{a_3}^{b_3} f(x, y, z) z dz}{\int_{a_3}^{b_3} f(x, y, z) dz} - \frac{\sum_{i=1}^n f(x, y, \xi_i) \xi_i \Delta z_i}{\sum_{i=1}^n f(x, y, \xi_i) \Delta z_i} \right| < \varepsilon$$

is uniformly true for all $(x, y) \in X \times Y$, where

$$\Delta z_i \triangleq z_i - z_{i-1} (i = 1, \cdots, n)$$

and ξ_i takes its value in $[z_{i-1}, z_i]$ arbitrarily. **Q.E.D.**

Theorem 2 Given a double-input single-output continuous stochastic system:

$$\underline{s} = S(X \times Y, Z, f(x, y, z)),$$

where $X = [a_1, b_1]$, $Y = [a_2, b_2]$ and $Z = [a_3, b_3]$ are finite real number intervals. If

$$(\forall(x, y) \in X \times Y) \left(\int_{a_3}^{b_3} f(x, y, z) dz > 0 \right),$$

then there exists a group of fuzzy inference rules:

$$\text{If } (x, y) \text{ is } D_i \text{ then } z \text{ is } C_i, i = 1, 2, \dots, n, \quad (20)$$

where $D_i \in \mathcal{F}(X \times Y)$ and $C_i \in \mathcal{F}(Z)$, such that the fuzzy system $\bar{s}$ constructed by the group of fuzzy inference rules can approximate the continuous stochastic system $\underline{s}$ to arbitrarily given precision.

Proof We make a partition of interval $Z = [a_3, b_3]$ as the following:

$$a_3 = z_0 < z_1 < \cdots < z_n = b_3.$$

Write $\Delta z_i = z_i - z_{i-1} (i = 1, 2, \cdots, n)$ and

$$\lambda = \max \{ \Delta z_i | i = 1, 2, \cdots, n \}.$$

Then we can get two Riemann sums

$$\sum_{i=1}^n f(x, y, z_i) z_i \Delta z_i, \quad \sum_{i=1}^n f(x, y, z_i) \Delta z_i.$$

From the condition of the theorem and Lemma 5, the following is true: $\exists \delta_1 > 0$, when $\lambda < \delta_1$, then

$$(\forall(x, y) \in X \times Y) \left(\sum_{i=1}^n f(x, y, z_i) \Delta z_i > 0 \right).$$

So we have

$$\begin{aligned}
\underline{s}(x, y) &= E(\zeta | \zeta = x, \eta = y) = \frac{\int_{a_3}^{b_3} f(x, y, z) z dz}{\int_{a_3}^{b_3} f(x, y, z) dz} \\
&\approx \frac{\sum_{i=1}^n f(x, y, z_i) z_i \Delta z_i}{\sum_{i=1}^n f(x, y, z_i) \Delta z_i} = \sum_{i=1}^n \left(\frac{f(x, y, z_i) \Delta z_i}{\sum_{j=1}^n f(x, y, z_j) \Delta z_j} \right) z_i \\
&= \sum_{i=1}^n \left(\frac{\mu_{D_i}(x, y) \Delta z_i}{\sum_{j=1}^n \mu_{D_i}(x, y) \Delta z_j} \right) z_i = \sum_{i=1}^n \mu_{D_i^*}(x, y) z_i,
\end{aligned}$$

where we have given a definition as follows:

$$\mu_{D_i}(x, y) \triangleq f(x, y, z_i) / M, \quad (21)$$

and $M \triangleq \max \{f(x, y, z) | (x, y, z) \in X \times Y \times Z\}$ and

$$\mu_{D_i^*}(x, y) \triangleq \frac{\mu_{D_i}(x, y) \Delta z_i}{\sum_{j=1}^n \mu_{D_j}(x, y) \Delta z_j}.$$

Clearly $D_i, D_i^* \in \mathcal{F}(X \times Y)$.

For any given approximation precision $\varepsilon > 0$, since $f(x, y, z)$ is continuous, from Lemma 6, $\exists \delta_2 > 0$ and $\delta_2 < \delta_1$, when $\lambda < \delta_2$, for all $(x, y) \in X \times Y$, that

$$\left| \underline{s}(x, y) - \sum_{i=1}^n \mu_{D_i^*}(x, y) z_i \right| < \frac{\varepsilon}{2}$$

holds uniformly.

By using the partition nodes $z_j (j = 0, 1, \dots, n)$, construct $n + 1$ fuzzy sets $C_j (j = 0, 1, \dots, n)$ on Z , which form a fuzzy partition of Z , i.e., make a kind of fuzziness of the nodes $z_j (j = 0, 1, \dots, n)$. Put $\mathcal{D} = \{D_i | 1 \leq i \leq n\}$ and $\mathcal{C} = \{C_i | 1 \leq i \leq n\}$. Regarding $\mathcal{D}$ and $\mathcal{C}$ as linguistic variables, a group of fuzzy inference rules can be formed as follows:

$$\text{If } (x, y) \text{ is } D_i \text{ then } z \text{ is } C_i, i = 1, 2, \dots, n. \quad (22)$$

We still use CRI method to make a fuzzy system $\bar{s}$. Firstly, from i -th fuzzy inference rule of (22), a fuzzy relation $R_i \triangleq D_i \times C_i$ on $X \times Y \times Z$

is formed, where its membership function is the following:

$$\mu_{R_i}(x, y, z) = \mu_{D_i}(x, y) \wedge \mu_{C_i}(z).$$

Then a whole fuzzy inference relation $R \triangleq \bigcup_{i=1}^n R_i$ is obtained as follows:

$$\begin{aligned}
\mu_R(x, y, z) &= \bigvee_{i=1}^n \mu_{R_i}(x, y, z) \\
&= \bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z))
\end{aligned}$$

For any a fuzzy set $D \in \mathcal{F}(X \times Y)$, a fuzzy inference result $C \in \mathcal{F}(Z)$ should be got by R , where $C \triangleq D \circ R$, i.e.,

$$\mu_C(z) = \bigvee_{(x, y) \in X \times Y} (\mu_D(x, y) \wedge \mu_R(x, y, z)) \quad (23)$$

For arbitrarily given an input $(x', y') \in X \times Y$, the point (x', y') should be turned into a fuzzy set:

$$\mu_{D'}(x, y) \triangleq \chi_{\{(x', y')\}}(x, y).$$

Substituting D' into (23), we obtain result of reasoning $C' \in \mathcal{F}(Z)$ as follows:

$$\mu_{C'}(z) = \mu_R(x', y', z) = \bigvee_{i=1}^n (\mu_{D_i}(x', y') \wedge \mu_{C_i}(z)).$$

It is easy to know that $\mu_{C'}(z) > 0$. Let

$$z' = \int_{a_3}^{b_3} z \mu_{C'}(z) dz / \int_{a_3}^{b_3} \mu_{C'}(z) dz.$$

And (x', y') is replaced by general point (x, y) in $X \times Y$ and z' by $\bar{s}(x, y)$. So we get a function as being $\bar{s} : X \times Y \rightarrow Z$ as follows:

$$\bar{s}(x, y) \triangleq \frac{\int_{a_3}^{b_3} \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z)) \right] z dz}{\int_{a_3}^{b_3} \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z)) \right] dz}, \quad (24)$$

which is still called a fuzzy approximation system of S , or called a double-input single fuzzy system, denoted by the following symbol:

$$\bar{s} = S(X \times Y, Z, \mathcal{D}, C).$$

Because for any a point $(x, y) \in X \times Y$,

$$\int_{a_3}^{b_3} \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z)) \right] dz > 0,$$

and by using lemma 5, $\exists \delta_3 > 0$, when $\lambda < \delta_3$, we have that, for $(x, y) \in X \times Y$,

$$\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z_j)) \right] \Delta z_j > 0.$$

Noticing the Riemann sum of (24) and C_i being with Kronecker property, we have the following equation:

$$\begin{aligned} \bar{s}(x, y) &\approx \frac{\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z_j)) \right] z_j \Delta z_j}{\sum_{j=1}^n \left[\bigvee_{i=1}^n (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z_j)) \right] \Delta z_j} \\ &= \frac{\sum_{j=1}^n \mu_{D_j}(x, y) z_j \Delta z_j}{\sum_{j=1}^n \mu_{D_j}(x, y) \Delta z_j} = \sum_{i=1}^n \frac{\mu_{D_i}(x, y) \Delta z_i}{\sum_{j=1}^n \mu_{D_j}(x, y) \Delta z_j} z_i \\ &= \sum_{i=1}^n \mu_{D_i^*}(x, y) z_i, \\ \mu_{D_i^*}(x, y) &\triangleq \frac{\mu_{D_i}(x, y) \Delta z_i}{\sum_{j=1}^n \mu_{D_j}(x, y) \Delta z_j} \end{aligned}$$

From Lemma 6, $\exists \delta_4 > 0$ and $\delta_4 < \delta_3$, when $\lambda < \delta_4$, for all $(x, y) \in X \times Y$, it uniformly holds that

$$\left| \bar{s}(x, y) - \sum_{i=1}^n \mu_{D_i^*}(x, y) z_i \right| < \frac{\varepsilon}{2}.$$

At last, we can take $\delta = \min \{\delta_2, \delta_4\}$. When $\lambda < \delta$, for all $(x, y) \in X \times Y$, $|\bar{s}(x) - \underline{s}(x)| < \varepsilon$ holds uniformly. This means that fuzzy system $\bar{s}$ can approximate continuous stochastic system $\underline{s}$ to arbitrarily given ε . **Q.E.D.**

Remark 1 In the proof of the theorem, fuzzy sets D_i are formed by (21). In fact, we can also make them as follows:

$$\mu_{D_i}(x, y) \triangleq \frac{f(x, y, z_i)}{\sum_{j=1}^n f(x, y, z_j)}, \quad (25)$$

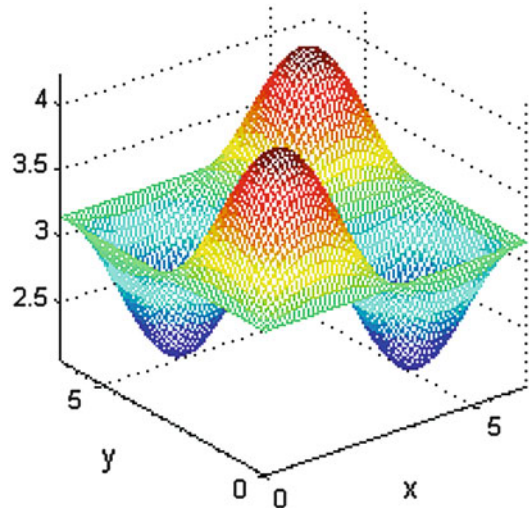
Based on (25), we can also prove the theorem. Here omits it. **Q.E.D.**

Example 2 Given a continuous stochastic system $\underline{s} = S(X \times Y, Z, f(x, y, z))$, where

$$\begin{aligned} X = Y = Z &= [0, 2\pi], \\ f(x, y, z) &= \frac{1 - \sin x \sin y \sin z}{8\pi^3} \end{aligned}$$

Based on Theorem 2, we have the following equation and its image refers to Fig. 10.

$$\begin{aligned} \underline{s}(x, y) &= E(\zeta | \zeta = x, \eta = y) \\ &= \frac{\int_0^{2\pi} f(x, y, z) z dz}{\int_0^{2\pi} f(x, y, z) dz} = \pi + \sin x \sin y. \end{aligned}$$

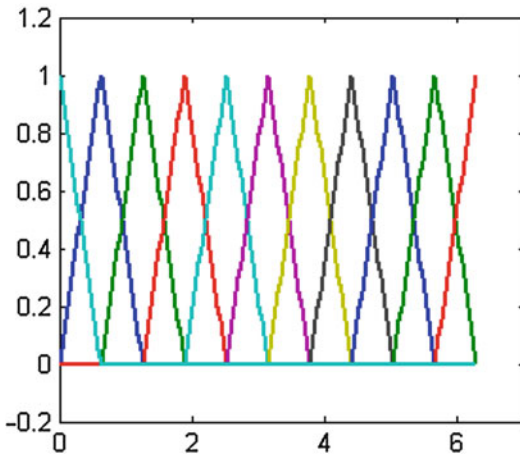


Fuzzy System Representations of Stochastic Systems, Fig. 10 Output surface of the system $\underline{s}$

Now we consider constructing a fuzzy approximation system $\bar{s}$. First make a partition of Z :

$$z_0 = 0, z_1 = 0.2\pi, z_2 = 0.4\pi, \dots, z_9 = 1.8\pi, z_{10} = 2\pi;$$

then fuzzy sets $C_j (j = 0, 1, \dots, 10)$ are formed as Fig. 11. The fuzzy sets $\mu_{D_i}(x, y)$ are defined as the following equation (see (21)) where $M = \frac{1}{4\pi^2}$, which images refer to Fig. 12.



Fuzzy System Representations of Stochastic Systems, Fig. 11 Membership function of fuzzy sets C_i

$$\begin{aligned} \mu_{D_i}(x, y) &= \frac{f(x, y, z_i)}{M} \\ &= \frac{1}{2} \left(1 - \sin x \sin y \sin \frac{i\pi}{5} \right) \end{aligned}$$

We calculate the following equation:

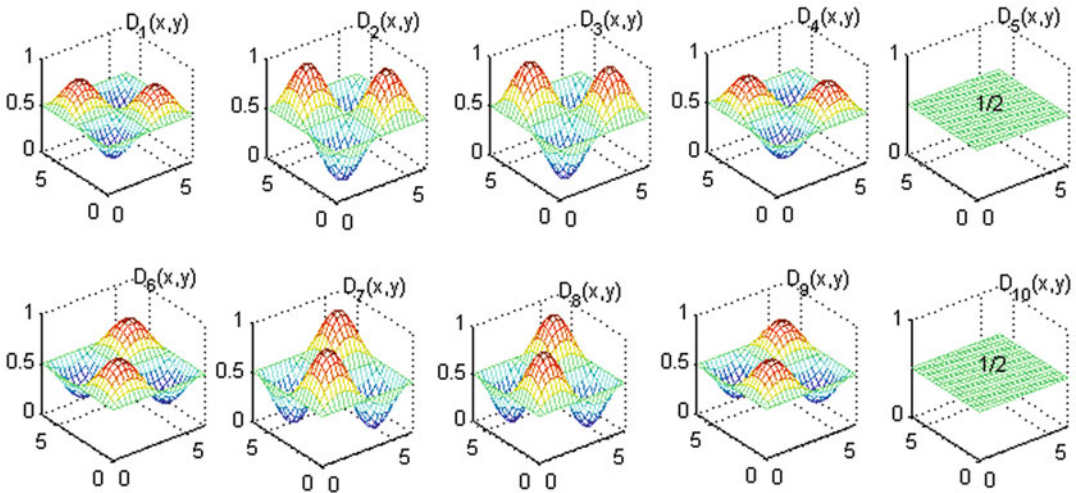
$$\bar{s}(x, y) = \frac{\int_0^{2\pi} \left[\bigvee_{i=1}^{10} (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z)) \right] z dz}{\int_0^{2\pi} \left[\bigvee_{i=1}^{10} (\mu_{D_i}(x, y) \wedge \mu_{C_i}(z)) \right] dz}$$

which image refers to Fig. 13 and the surface of the error $\bar{s}(x, y) - \underline{s}(x, y)$ refers to Fig. 14.

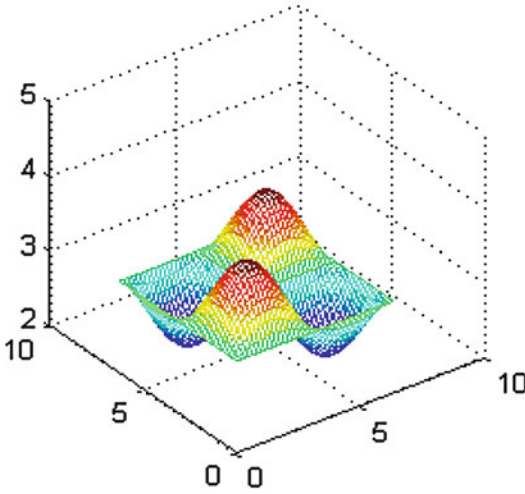
If the partition nodes $\{z_j\}$ are increased of double number, i.e.,

$$z_0 = 0, z_1 = 0.1\pi, z_2 = 0.2\pi, \dots, z_{20} = 2\pi,$$

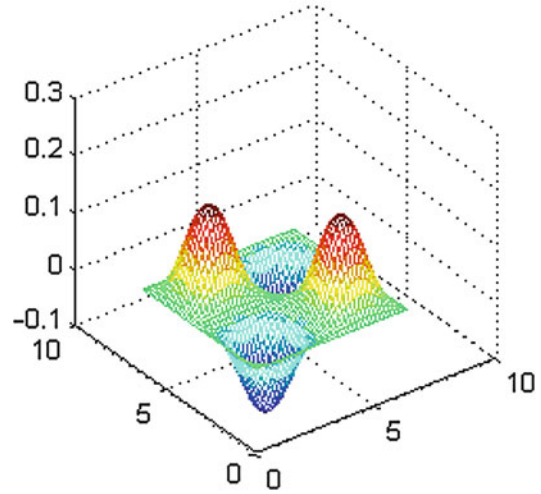
then the approximation precision increases better, where the error surface of error function $\bar{s}(x, y) - \underline{s}(x, y)$ refers to Fig. 15. And if the partition nodes $\{z_j\}$ are increased more, then we can see that the error between $\bar{s}(x, y)$ and $\underline{s}(x, y)$ is very small from Fig. 16.



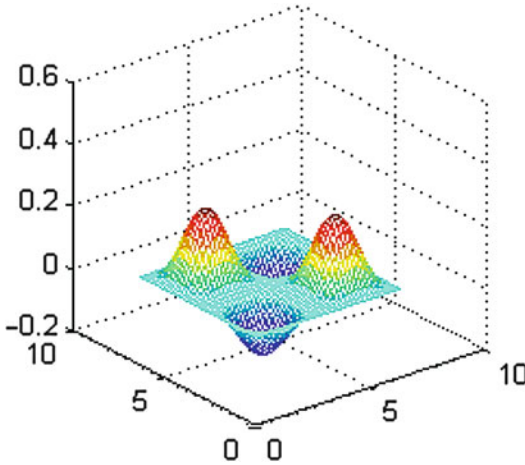
Fuzzy System Representations of Stochastic Systems, Fig. 12 Membership function surfaces of fuzzy sets D_i



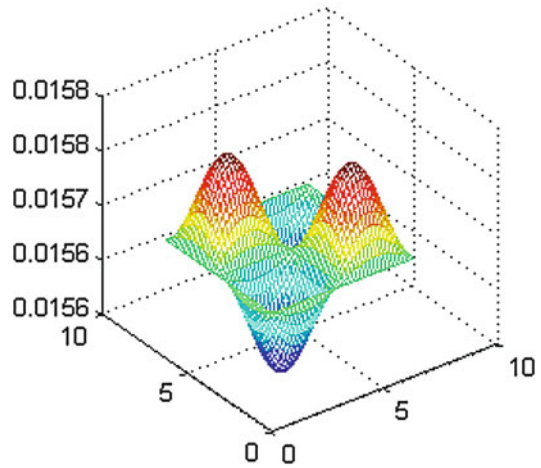
Fuzzy System Representations of Stochastic Systems, Fig. 13 Output surface of fuzzy system $\bar{s}$



Fuzzy System Representations of Stochastic Systems, Fig. 15 The error surface when $\lambda = 0.1\pi$



Fuzzy System Representations of Stochastic Systems, Fig. 14 The error surface when $\lambda = 0.2\pi$



Fuzzy System Representations of Stochastic Systems, Fig. 16 The error surface when $\lambda = 0.01\pi$

Fuzzy Reasoning Representations of Discrete Stochastic Systems

First consider single-input single-output uncertainty system $s = S(X, Y)$. Suppose that we have known some probability information about the system S . And we should be seeing about the system from stochastic viewpoint.

Let X and Y be measurable sets in real number space $\mathbb{R}$, and input random variable ξ and output random variable η be defined in probability spaces

$(X, \mathcal{B}_1, P_1)$ and $(Y, \mathcal{B}_2, P_2)$, respectively, where $\mathcal{B}_1$ and $\mathcal{B}_2$ are Borel σ -fields on X and Y respectively, and P_1 and P_2 be probability measures on $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. As done in Introduction, we can construct a joint probability space $(\Omega, \mathcal{F}, P)$. Then (ξ, η) is a random vector on the probability space $(\Omega, \mathcal{F}, P)$.

Suppose that we have mastered the discrete probability distribution $\{P(x_i, y_j) | 1 \leq i \leq n, 1 \leq j \leq m\}$ of the system, where $X = [a_1, b_1]$, $Y = [a_2, b_2]$, and

$$a_1 \leq x_1 < \cdots < x_n \leq b_1, \\ y_0 \triangleq a_2 < y_1 < \cdots < y_m = b_2$$

Assume

$$(\forall i \in \{1, 2, \dots, n\}) \left(\sum_{j=1}^m P(x_i, y_j) > 0 \right). \text{ Write}$$

$$\underline{s}(x_i) \triangleq E(\eta | \xi = x_i) = \frac{\sum_{j=1}^m P(x_i, y_j) y_j}{\sum_{j=1}^m P(x_i, y_j)}, \quad (26)$$

This is regarded as the response of S after x_i is input. So, if we write that

$$\underline{s} = S(X, Y \{P(x_i, y_j) | 1 \leq i \leq n, 1 \leq j \leq m\}),$$

then $\underline{s}$ defined as (26) is called a discrete stochastic approximation system of the uncertainty S or simply called a discrete stochastic system.

Theorem 3 Given arbitrarily a discrete stochastic system as follows

$$\underline{s} = S(X, Y \{P(x_i, y_j) | 1 \leq i \leq n, 1 \leq j \leq m\})$$

where $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are finite real number intervals, then there exists a group of fuzzy inference rules:

$$\text{If } x \text{ is } A_j \text{ then } y \text{ is } B_j, j = 1, 2, \dots, m, \quad (27)$$

where $A_j \in \mathcal{F}(X)$ and $B_j \in \mathcal{F}(Y)$, such that the fuzzy system $\bar{s}$ constructed by the group of fuzzy inference rules can approximate the discrete stochastic system $\underline{s}$ to some precision $\varepsilon > 0$, i.e.,

$$(\forall i \in \{1, 2, \dots, n\}) (|\bar{s}(x_i) - \underline{s}(x_i)| \leq \varepsilon),$$

and the smaller is $\lambda = \max_{1 \leq j \leq m} \{\Delta y_j\}$, the higher is precision, where $\Delta y_j = y_j - y_{j-1}, j = 1, 2, \dots, m$.

Proof First of all, we consider how to make a group of fuzzy inference rules (27). One side, $B_j \in \mathcal{F}(Y)$ is easy to get. In fact, just as in

Fig. 3, we can form “triangle waves” membership functions of the fuzzy sets, where subscript n should be replaced by m . So the linguistic variable $\mathcal{B} = \{B_j | 1 \leq j \leq m\}$ is obtained. Other side, we have to make fuzzy sets $A_j \in \mathcal{F}(X)$.

For every j , we construct a group of fuzzy sets:

$$\alpha_i \in \mathcal{F}(X), i = 1, 2, \dots, n,$$

as base functions as follows:

$$\mu_{\alpha_1}(x) = \begin{cases} (x - x_2)/(x_1 - x_2), & x_1 \leq x \leq x_2, \\ 0, & \text{otherwise;} \end{cases}$$

$$\mu_{\alpha_i}(x) = \begin{cases} (x - x_{i-1})/(x_i - x_{i-1}), & x_{i-1} \leq x \leq x_i, \\ (x - x_{i+1})/(x_i - x_{i+1}), & x_i < x \leq x_{i+1}, \\ 0, & \text{otherwise,} \end{cases}$$

$$i = 2, 3, \dots, n-1;$$

$$\mu_{\alpha_n}(x) = \begin{cases} (x - x_{n-1})/(x_n - x_{n-1}), & x_{n-1} \leq x \leq x_n, \\ 0, & \text{otherwise.} \end{cases}$$

Second, (26) is turned into the following:

$$\underline{s}(x_i) = \frac{\sum_{j=1}^m P(x_i, y_j) y_j}{\sum_{j=1}^m P(x_i, y_j)}$$

$$= \sum_{j=1}^m \frac{P(x_i, y_j)}{\sum_{j=1}^m P(x_i, y_j)} y_j = \sum_{j=1}^m a_{ij} y_j,$$

where

$$a_{ij} \triangleq \frac{P(x_i, y_j)}{\sum_{j=1}^m P(x_i, y_j)}.$$

Then by means of the above, the weighted mean of fuzzy sets α_i , we form fuzzy sets A_j as follows:

$$\mu_{A_j}(x) \triangleq \sum_{i=1}^n a_{ij} \mu_{\alpha_i}(x), j = 1, 2, \dots, m, \quad (28)$$

where the group of weight vector as follows

$$\{(a_{1j}, a_{2j}, \dots, a_{nj}) | 1 \leq j \leq m\}$$

will be determined. Thus another linguistic variable as follows

$$\mathcal{A} = \{A_j | 1 \leq j \leq m\}$$

is regarded as being obtained, and we have got a group of fuzzy inference rules as (27) as follows:

If x is A_j then y is $B_j, j = 1, 2, \dots, m$.

As in the proof of theorem 2, by CRI method, we can make a fuzzy system $\bar{s}$:

$$\bar{s}(x) \triangleq \frac{\int_{a_2}^{b_2} \left[\bigvee_{k=1}^m (\mu_{A_k}(x) \wedge \mu_{B_k}(y_j)) \right] y dy}{\int_{a_2}^{b_2} \left[\bigvee_{k=1}^m (\mu_{A_k}(x) \wedge \mu_{B_k}(y_j)) \right] dy}. \quad (29)$$

This means that we get a fuzzy approximation system $\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$ of the uncertainty system S . Noticing the Riemann sum of (13.4) and B_j with Kronecker property, we have

$$\begin{aligned} \bar{s}(x) &\approx \frac{\sum_{j=1}^m \left[\bigvee_{k=1}^m (\mu_{A_k}(x) \wedge \mu_{B_k}(y_j)) \right] y_j \Delta y_j}{\sum_{j=1}^m \left[\bigvee_{k=1}^m (\mu_{A_k}(x) \wedge \mu_{B_k}(y_j)) \right] \Delta y_j} \\ &= \frac{\sum_{j=1}^m \mu_{A_j}(x) y_j \Delta y_j}{\sum_{j=1}^m \mu_{A_j}(x) \Delta y_j} \end{aligned}$$

Since $\mu_{A_j}(x_i) = a_{ij}$, we have

$$\bar{s}(x_i) \approx \frac{\sum_{j=1}^m \mu_{A_j}(x) y_j \Delta y_j}{\sum_{j=1}^m \mu_{A_j}(x) \Delta y_j} = \frac{\sum_{j=1}^m (a_{ij} \Delta y_j) y_j}{\sum_{j=1}^m a_{ij} \Delta y_j}. \quad (30)$$

After comparing (26) and (30), take

$$a_{ij} = P(x_i, y_j) M / \Delta y_j,$$

where $M \triangleq \min_{(i,j)} \left\{ \Delta y_j / P(x_i, y_j) \right\}$. Let

$$\varepsilon \triangleq \max \{ |\bar{s}(x_i) - \underline{s}(x_i)| | i = 1, 2, \dots, n \}.$$

Then we at last get the following result:

$$(\forall i \in \{1, 2, \dots, n\}) (|\bar{s}(x_i) - \underline{s}(x_i)| \leq \varepsilon).$$

Besides, for any given $\varepsilon > 0$, from (30), $\exists \delta > 0$, when $\lambda > \delta$, for all $x_i (i = 1, 2, \dots, n)$, we have the result:

$$|\bar{s}(x_i) - \underline{s}(x_i)| < \varepsilon.$$

Clearly, the more is λ , the higher is precision.

Q.E.D.

We now consider double-input single-output uncertainty system $s = S(X \times Y, Z)$. Let X, Y and Z be measurable set in real number space $\mathbb{R}$ and input random variables ξ and η and output random variable ζ be defined, respectively, in probability spaces $(X, \mathcal{B}_1, P_1)$, and $(Y, \mathcal{B}_2, P_2)$, where $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$ are Borel σ -fields on X, Y , and Z , and P_1, P_2 , and P_3 are probability measures on $\mathcal{B}_1, \mathcal{B}_2$, and $\mathcal{B}_3$, respectively. We can also get the joint probability space $(\Omega, \mathcal{F}, P)$ in the same. So (ξ, η, ζ) becomes a random vector on $(\Omega, \mathcal{F}, P)$. Suppose that we have known a discrete probability distribution:

$$\bar{P} \triangleq \left\{ P(x_i, y_j, z_k) | 1 \leq i \leq n, 1 \leq j \leq m, 1 \leq k \leq p \right\},$$

where $X = [a_1, b_1], Y = [a_2, b_2], Z = [a_2, b_2]$, and

$$\begin{aligned} a_1 &\leq x_1 < \dots < x_n \leq b_1, \\ a_2 &\leq y_1 < \dots < y_m \leq b_2, \\ z_0 &\triangleq a_3 < z_1 < \dots < z_p = b_3 \end{aligned}$$

Assume that $(\forall i, j) \left(\sum_{k=1}^p P(x_i, y_j, z_k) > 0 \right)$.

Write

$$\begin{aligned} \underline{s}(x_i, y_j) &\triangleq E(\zeta | \xi = x_i, \eta = y_j) \\ &= \frac{\sum_{k=1}^p P(x_i, y_j, z_k) z_k}{\sum_{k=1}^p P(x_i, y_j, z_k)} \end{aligned} \quad (31)$$

It is regarded as a response after an input (x_i, y_j) is input into the system S . Put

$$\underline{s} = S(X \times Y, Z, \bar{P}) \quad (32)$$

Then $\underline{s}$ defined by (32) is called a double-input single-output discrete stochastic approximation system of S or simply a double-input single-output discrete stochastic system.

Theorem 4 Given a discrete stochastic system from (32), where $X = [a_1, b_1]$, $Y = [a_2, b_2]$, $Z = [a_2, b_2]$ are finite real number intervals, then there exists a group of fuzzy inference rules:

If (x, y) is D_k then z is C_k , $k = 1, 2, \dots, p$, (33) where $D_k \in \mathcal{F}(X \times Y)$ and $C_k \in \mathcal{F}(Z)$, such that the fuzzy $\bar{s}$ system constructed by the group of fuzzy inference rules can approximate the discrete stochastic system $\underline{s}$ to some precision $\varepsilon > 0$, i.e., for any $a_i \in \{1, 2, \dots, n\}$ and for any $a_j \in \{1, 2, \dots, m\}$, we have

$$|\bar{s}(x_i, y_j) - \underline{s}(x_i, y_j)| \leq \varepsilon$$

And the smaller is $\lambda = \max_{1 \leq k \leq p} \{\Delta z_k\}$, the higher is precision, where $\Delta z_k = z_k - z_{k-1}$ ($k = 1, 2, \dots, p$).

The proof is the same as one of theorem 3, and we omit it. **Q.E.D.**

Reducibility in the Transformations Between Fuzzy Systems and Stochastic Systems

Take single-input and single-output open-loop system as being $s = S(X, Y)$ as an example to

discuss the problem on reducibility in the transformations between fuzzy systems and stochastic systems. We only consider continuous systems as discrete systems are special cases of continuous systems and can be treated with no difficulty. Here $X = [a_1, b_1]$ and $Y = [a_2, b_2]$ are finite real number intervals.

First, suppose that we have known a fuzzy system as being $\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$, where $\mathcal{A} = \{A_i | 1 \leq i \leq n\}$ and $\mathcal{B} = \{B_i | 1 \leq i \leq n\}$ are fuzzy partitions of X and Y , respectively. $\mathcal{A}$ and $\mathcal{B}$ are regarded as linguistic variables, and we can get a group of fuzzy inference rules:

If x is A_i then y is B_i , $i = 1, 2, \dots, n$.

For convenience, the group of fuzzy inference rules is simply denoted by the following:

$$\mathcal{A} \rightarrow \mathcal{B}. \quad (33)$$

By (33) we have the input-output function of the fuzzy system $\bar{s}$:

$$\bar{s}(x) = \frac{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right] dy}{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right] dy}, \quad (34)$$

where θ is a fuzzy implication operator with the condition:

$$(\forall (a, b) \in [0, 1]) (\theta(a, 1) = a, \theta(a, \theta) = 0) \quad (35)$$

From above discuss, there must exist a joint probability space $(\Omega, \mathcal{F}, P)$, where $\Omega = X \times Y$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ and $P = P_1 \times P_2$, and random variables ξ and η are defined in probability spaces $(X, \mathcal{F}_1, P_1)$ and $(Y, \mathcal{F}_2, P_2)$. After redefining ξ and η in $(\Omega, \mathcal{F}, P)$, a random vector (ξ, η) is obtained, which obeys the probability distribution based on the following probability density:

$$f(x, y) = \bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) / H(2, n, \theta, \vee),$$

$$H(2, n, \theta, \vee) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \left[\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right] dx dy \quad (36)$$

By (5), we can get a stochastic system as follows:

$$\underline{s} = S(X, Y, f(x, y))$$

which input–output function is as the following:

$$\underline{s}(x) = \int_{a_2}^{b_2} f(x, y) y dy / \int_{a_2}^{b_2} f(x, y) dy.$$

It is easy to know that the above equation is the same as Eq. (34), i.e., $\bar{s}(x) \equiv \underline{s}(x)$. Now we show that $f(x, y)$ defined by (36) can be returned into the original fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$. In fact, by using original partition nodes $x_1 < x_2 < \dots < x_n$ and $y_1 < y_2 < \dots < y_n$ of X and Y , respectively, we get a fuzzy inference rule group, denoted by the following symbol:

$$\mathcal{A}' = \{A'_i | 1 \leq i \leq n\} \rightarrow \mathcal{B}' = \{B'_i | 1 \leq i \leq n\}.$$

First we put $\mu_{A'_i}(x) \triangleq f(x, y_i) / M$, where

$$M \triangleq \max \{ f(x, y) | (x, y) \in X \times Y \}.$$

Noticing that

$$\begin{aligned} M &= \max \{ f(x, y) | (x, y) \in X \times Y \} \\ &= \frac{\max_{(x, y) \in X \times Y} \left\{ \bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right\}}{H(2, n, \theta, \vee)} \\ &= \frac{1}{H(2, n, \theta, \vee)} \end{aligned}$$

we have that $M \cdot H(2, n, \theta, \vee) = 1$. It is easy to verify that, for any $i \in \{1, 2, \dots, n\}$,

$$\begin{aligned} \mu_{A'_i}(x) &= \frac{f(x, y_i)}{M} \\ &= \frac{\bigvee_{j=1}^n \theta(\mu_{A_j}(x), \mu_{B_j}(y_i))}{MH(2, n, \theta, \vee)} = \mu_{A_i}(x) \end{aligned}$$

So $\mathcal{A}' = \mathcal{A}$. Constructing $\mathcal{B}' = \{B'_i | 1 \leq i \leq n\}$ is simple and of some freedom since we only demand that they are a kind of fuzzification of the peak points $y_i (i = 1, 2, \dots, n)$ and a partition of Y . Thus, we can directly take

$$B'_i = B_i, i = 1, 2, \dots, n.$$

Then we have $\mathcal{B}' = \mathcal{B}$. Thus we have reverted $f(x, y)$ into original fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$. This is one side reducibility.

Now we consider another side reducibility. Given a stochastic system $\underline{s} = S(X, Y, f(x, y))$. We know that its input–output function is as

$$\underline{s}(x) = \int_{a_2}^{b_2} f(x, y) y dy / \int_{a_2}^{b_2} f(x, y) dy.$$

By theorem 1, there exists a group of fuzzy inference rules $\mathcal{A} \rightarrow \mathcal{B}$ such that the fuzzy system as follows

$$\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$$

constructed by $\mathcal{A} \rightarrow \mathcal{B}$ can approximate the stochastic system $\underline{s} = S(X, Y, f(x, y))$ to a given precision, i.e., $\bar{s}(x) \approx \underline{s}(x)$ (Notice that is not $\bar{s}(x) \equiv \underline{s}(x)$, in other words,

$$(\forall x \in X) \left(\lim_{\lambda \rightarrow 0} \bar{s}(x) = \underline{s}(x) \right), \quad (37)$$

where $\lambda = \max \{ \Delta y_j | j = 1, 2, \dots, n \}$. These A_i are made by using (9) or (3.9) and B_i can be taken as triangle waves membership functions referring to Fig. 2. Then the input–output function of the fuzzy system $\bar{s}$, which is more general than one in (12), is as follows:

$$\bar{s}(x) = \frac{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right] y dy}{\int_{a_2}^{b_2} \left[\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y)) \right] dy},$$

where the fuzzy implication operator θ still needs to meet the condition (35). So we get a stochastic system as follows:

$$\underline{s} = S(X, Y, f'(x, y)),$$

where

$$f'(x, y) = \frac{\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y))}{H(2, n, \theta, \vee)}.$$

The meaning of $H(2, n, \theta, \vee)$ is the same as before. We should consider two cases for the reducibility.

Case 1 By (3.9), stipulate $A_i(x) \triangleq f(x, y_i)/M$, then at the division points $y_j (j = 1, \dots, n)$ of Y , we have

$$\begin{aligned} f'(x, y_j) &= \frac{\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y_j))}{H(2, n, \theta, \vee)} \\ &= \frac{\mu_{A_i}(x)}{H(2, n, \theta, \vee)} = \frac{f(x, y_j)}{H(2, n, \theta, \vee)M} \\ &= \alpha(n) f(x, y_j), \\ \alpha(n) &\triangleq \frac{1}{H(2, n, \theta, \vee)M} \end{aligned} \quad (38)$$

For given a partition on Y , n is fixed, so $\alpha(n)$ is constant. Equation (38) means that at every nodal point y_j , $f'(x, y_j)$ is reverted to $f(x, y_j)$ under ignoring a constant factor $\alpha(n)$. Because h can be decreased arbitrarily, $f'(x, y)$ should be regarded as having been reverted to $f(x, y)$ under ignoring a constant factor.

Case 2 Suppose that nodes $y_j (j = 1, 2, \dots, n)$ of Y are equidistant, i.e., $\Delta y_j = \lambda (j = 1, 2, \dots, n)$. By (36), we stipulate that

$$\mu_{A_i}(x) \triangleq \frac{f(x, y)}{\sum_{j=1}^n f(x, y_j)},$$

then at the nodal points $y_j (j = 1, 2, \dots, n)$ of Y , we have

$$\begin{aligned} f'(x, y_j) &= \frac{\bigvee_{i=1}^n \theta(\mu_{A_i}(x), \mu_{B_i}(y_j))}{H(2, n, \theta, \vee)} \\ &= \frac{\mu_{A_i}(x)}{H(2, n, \theta, \vee)} = \frac{f(x, y_j)}{H(2, n, \theta, \vee) \sum_{i=1}^n f(x, y_j)} \\ &= \frac{\lambda}{H(2, n, \theta, \vee)} \cdot \frac{f(x, y_j)}{\sum_{i=1}^n f(x, y_i) \lambda} \\ &\approx \beta(\lambda) \frac{f(x, y_j)}{\int_Y f(x, y) dy} = \beta(\lambda) \frac{f(x, y_j)}{f_\xi(x)} \\ &= \beta(h) f_{\eta|\xi=x}(y_j|x) \end{aligned} \quad (39)$$

where $\beta(\lambda) \triangleq \lambda/H(2, n, \theta, \vee)$, which is constant under given a partition on Y , $f_\xi(x) \triangleq \int_Y f(x, y) dy$ is marginal probability density, and $f_{\eta|\xi=x}(y|x) \triangleq f(x, y)/f_\xi(x)$ is conditional probability density. Based on the above equation, $f'(x, y)$ should be regarded as having been reverted to the conditional probability density $f_{\eta|\xi=x}(y|x)$ under ignoring a constant factor.

Remark 2 In $\beta(\lambda) \triangleq \lambda/H(2, n, \theta, \vee)$, it is easy to learn that two parameters λ and n are correlated to each other. So $\beta(\lambda)$ can also be written as $\beta(\lambda)$, but not written as $\beta(\lambda, n)$. **Q.E.D.**

Remark 3 Noticing that $f'(x, y)$ is a continuous function, from numerical analysis and based on (38), we can easily learn that $f'(x, y)$ can be regarded as a interpolation function with the node group

$$\left\{ \left(y_j, \alpha(n)f(x, y_j) \right) \mid j = 1, 2, \dots, n \right\}.$$

At every node y_j , $f'(x, y)$ is strictly equal to the nodal function value $\alpha(n)f(x, y_j)$. There is similar understanding about (39). **Q.E.D.**

The Uncertainty Systems with One-Dimensional Random Variable and Their Representations

From the above sections in this entry, we can find a situation that the probability density function with respect to uncertainty systems are at least two-dimensional, i.e., the random vectors we dealt with are at least two-dimensional, say (ξ, η) where ξ is in essence defined on input universe X and η on output universe Y . This is not strange because an uncertainty system is at least of one input variable $x \in X$ and one output variable $y \in Y$. However, when we learn probability theory and its applications, random variables with respect to a lot of random experiments are one-dimensional, say ξ , and if it has probability density function, it is a one-dimensional function, say $f(x)$.

Naturally we should ask such a question: What kind of uncertainty systems are only with respect to one-dimensional random variables, or one-dimensional probability density functions? We can guess that such uncertainty systems may have some special characteristics or trivialities. And such uncertainty systems may have many cases. We can only discuss several typical cases.

Typical Case 1: Pure Certainty Systems

As a Cantor's set can be regarded as a fuzzy set, certainty systems can also be regarded as special uncertainty systems. Now we consider a special open-loop system $s = S(X, Y)$, where $X = [a, b]$ and that $Y = \{y_0\}$ is a singleton. We have known that the symbol s has double meanings, i.e., it not only abstractly represents the system, but also means the relation between input and output of the system,

$$s : X \rightarrow Y, \quad x \mapsto y = s(x) \triangleq y_0.$$

Since $(\forall x \in X)(s(x) = y_0)$, this is a special certainty system, and also a trivial system, and its output is a step function.

We will use CRI method to construct its fuzzy approximation system. First make a partition of X :

$$a = x_1 < x_2 < \dots < x_n = b.$$

Then the nodes $x_i (i = 1, 2, \dots, n)$ are fuzzified as Fig. 2 to obtain fuzzy sets $A_i \in \mathcal{F}(X) (i = 1, 2, \dots, n)$ (note that $n + 1$ subscripts $0, 1, \dots, n$ should be changed to n subscripts $1, 2, \dots, n$). So a fuzzy partition of X is got as $\mathcal{A} = \{A_i \mid i = 1, 2, \dots, n\}$. The fuzzy sets on Y are easily formed as

$$B_i \triangleq Y = \{y_0\}, i = 1, 2, \dots, n.$$

Thus a fuzzy partition of Y is also got as follows:

$$\mathcal{B} = \{B_i \mid i = 1, 2, \dots, n\}.$$

This means that we have a fuzzy inference rule group: $\mathcal{A} \rightarrow \mathcal{B}$. Therefore a fuzzy approximation system is constructed as

$$\bar{s} = S(X = [a, b], Y = \{y_0\}, \mathcal{A}, \mathcal{B}).$$

For simplification, fuzzy implication operator θ is taken as $\wedge$. We have

$$\begin{aligned} p(x, y) &= p(x, y_0) \\ &= \bigvee_{i=1}^n [\mu_{A_i}(x) \wedge \mu_{B_i}(y_0)] = \bigvee_{i=1}^n \mu_{A_i}(x) \end{aligned}$$

Then the input-output relation should be

$$\bar{s}(x) = \int_Y y p(x, y_0) dy / \int_Y p(x, y_0) dy.$$

Let

$$H(1, n, \wedge, \vee) \triangleq \int_X p(x, y_0) dx$$

and it is easy to know that $\int_X p(x, y_0) dx > 0$. So we can put

$$\begin{aligned} f(x) &\triangleq f(x, y_0) \\ &= \frac{p(x, y_0)}{H(1, n, \wedge, \vee)} = \frac{\bigvee_{i=1}^n \mu_{A_i}(x)}{H(1, n, \wedge, \vee)} \end{aligned} \quad (40)$$

So there should exist a random variable as follows:

$$\xi : X \rightarrow \mathbb{R}, x \mapsto \xi(x)$$

defined on the probability space $(X, \mathcal{F}, P)$, to obey a probability distribution with probability density function $f(x)$, where $\mathcal{F}$ is a Borel σ -field on X . If we let $\Omega \triangleq X \times \{y_0\}$, then $\mathcal{F}$ can be regarded as a Borel σ -field on Ω . And ξ can be regarded as a random variable defined on Ω , i.e., without changing symbol redefined as follows:

$$\begin{aligned} \xi : \Omega &\rightarrow \mathbb{R} \\ \omega = (x, y_0) &\mapsto \xi(\omega) = \xi(x, y_0) \triangleq \xi(x), \end{aligned}$$

and P is also regarded as a probability on $(\Omega, \mathcal{F}, P)$. Thus we get a stochastic system

$$\underline{s} = S(X, Y, f(x, y_0)),$$

and its input-output relation is

$$\underline{s} = \int_Y y p(x, y_0) dy / \int_Y p(x, y_0) dy.$$

Clearly, $(\forall x \in X)(\bar{s}(x) = \underline{s}(x))$, which is consistent with the result in [Reducibility in the Transformations Between Fuzzy Systems and Stochastic Systems](#).

But is only a formal representation, because Y is a singleton and its measure is zero, and so for any $x \in X$,

$$\int_Y y p(x, y_0) dy = 0, \int_Y p(x, y_0) dy = 0,$$

which means that

$$\bar{s}(x) = \frac{\int_Y y p(x, y_0) dy}{\int_Y p(x, y_0) dy} = \frac{0}{0}.$$

Thus $\int_Y y p(x, y_0) dy / \int_Y p(x, y_0) dy$ is meaningless. It is not difficult to deal with. In fact, for any $\varepsilon > 0$, it is easy to know that $(\forall x \in X)(p(x, y_0) > 0)$, and so, for any $x \in X$,

$$\begin{aligned} \int_{y_0}^{y_0+\varepsilon} p(x, y_0) dy &= p(x, y_0) \int_{y_0}^{y_0+\varepsilon} dy \\ &= p(x, y_0) \varepsilon > 0 \end{aligned}$$

Thus, for any a point $x \in X$, we also have the following expression:

$$\begin{aligned} \bar{s}(x) &= \frac{\int_Y y p(x, y_0) dy}{\int_Y p(x, y_0) dy} = \frac{\int_{\{y_0\}} y p(x, y_0) dy}{\int_{\{y_0\}} p(x, y_0) dy} \\ &\triangleq \lim_{\varepsilon \rightarrow 0} \frac{\int_{y_0}^{y_0+\varepsilon} y p(x, y_0) dy}{\int_{y_0}^{y_0+\varepsilon} p(x, y_0) dy} = \lim_{\varepsilon \rightarrow 0} \frac{p(x, y_0) \int_{y_0}^{y_0+\varepsilon} y dy}{p(x, y_0) \varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \left(y_0 + \frac{\varepsilon}{2} \right) = y_0. \end{aligned}$$

Besides, clearly, $(\forall x \in X)(\bar{s}(x) = s(x))$ which means that, in this case, the fuzzy system $\bar{s}$ constructed by CRI method can accurately approximate the system s .

Example 3 Consider an illumination system. For simplification, suppose that there is only one lamp in the system. Usually, we open the lamp in evening, and suppose that the time of opening the lamp is between a and b . Let $X = [a, b]$ that is regarded as the input universe of the system. If we take $x \in X$, then it means that the lamp is opened at time x , and the lamp emits light, which is denoted by a symbol, say 1. Naturally we can take $Y = \{y_0\} = \{1\}$ as the output universe. Clearly, the input-output relation of the system is

$$s : X \rightarrow Y, x \mapsto y = s(x) \triangleq y_0 = 1.$$

This is a pure certain system, of course. We have a reason to ask: Now that this is a certain system, why does there exist a random variable ξ and a probability density function $f(x)$ that ξ

obeys? In fact, if we focus our attention on “what time we should open the lamp in evening,” this becomes a stochastic problem. And this is not contrary with the certain input–output relation

$$s : X \rightarrow Y, x \mapsto y = s(x) \triangleq y_0 = 1.$$

You know, “what time we should open the lamp in evening” depends on many factors, such as different area leads different opening lamp time due to time difference. By using random experiment, we should know that opening lamp is round about at several time, such as about 5 o’clock, about 6 o’clock, about 7 o’clock, etc. Generally suppose that the lamp is usually opened at about x_1 , about x_2 , $\dots$, about x_n , and denote $a = x_1$ and $b = x_n$. If $x_i (i = 1, 2, \dots, n)$ are fuzzified to get fuzzy sets $A_i \in \mathcal{F}(X) (i = 1, 2, \dots, n)$. Then we make fuzzy sets on Y as the following

$$\begin{aligned} B_i &\triangleq Y = \{y_0\}, \\ \mu_{B_i}(y) &= \chi_Y(y) = \chi_{\{y_0\}}(y) \\ i &= 1, 2, \dots, n, \end{aligned}$$

and $\mathcal{B} = \{B_i | i = 1, 2, \dots, n\}$ is formed. Thus we have a fuzzy inference rule group: $\mathcal{A} \rightarrow \mathcal{B}$, which means that we get a fuzzy system $\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$. This comes back to the way of obtaining (40), and the rest is clear. **Q.E.D.**

Typical Case 2: Pure Stochastic Systems

Consider another special open-loop system $s = S(X, Y)$, where $X \triangleq \{x_0\}$ is a singleton and $Y = [a, b]$, and its input–output relation should be

$$s : X \rightarrow Y, x_0 \mapsto y_0 = s(x_0).$$

Due to the uncertainty of the system, for the input x_0 , we do not in advance know which y_0 in Y should correspond to x_0 . So $s = S(X, Y)$ is a pure stochastic system. After x_0 is input, by using statistic method there are several output to correspond to it: about y_1 , about y_2 , $\dots$, about y_n . After $y_0, y_1, \dots, y_n$ are fuzzified, we get fuzzy sets

$$B_i \in \mathcal{F}(Y), i = 0, 1, \dots, n,$$

which are demanded to form a fuzzy partition of Y . Let

$$\begin{aligned} A_i &\triangleq X = \{x_0\}, \\ \mu_{A_i}(y) &= \chi_X(x) = \chi_{\{x_0\}}(x), \\ i &= 1, 2, \dots, n, \end{aligned}$$

and denote $\mathcal{A} = \{A_i | i = 1, 2, \dots, n\}$ and

$$\mathcal{B} = \{B_i | i = 1, 2, \dots, n\}$$

(not use B_0). Then a fuzzy inference rule group is got as $\mathcal{A} \rightarrow \mathcal{B}$. So we obtain a fuzzy system:

$$\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B}).$$

By CRI method, we have that

$$\begin{aligned} p(x, y) &= p(x_0, y) \\ &= \bigvee_{i=1}^n [\mu_{A_i}(x_0) \wedge \mu_{B_i}(y)] = \bigvee_{i=1}^n \mu_{B_i}(y) \end{aligned}$$

Since $(\forall y \in Y)(p(x_0, y) > 0)$,

$$H(1, n, \wedge, \vee) \triangleq \int_Y p(x_0, y) dy > 0.$$

Let

$$\begin{aligned} g(y) &\triangleq g(x_0, y) \\ &= \frac{p(x_0, y)}{H(1, n, \wedge, \vee)} = \frac{\bigvee_{i=1}^n \mu_{B_i}(y)}{H(1, n, \wedge, \vee)} \end{aligned} \quad (41)$$

Noticing that $\int_Y g(x_0, y) dy = 1$, the input–output relation should be as the following:

$$\begin{aligned} \bar{s}(x) &= \frac{\int_Y y g(x_0, y) dy}{\int_Y g(x_0, y) dy} = \int_Y y g(x_0, y) dy \\ &= \frac{\int_a^b y \left[\bigvee_{i=1}^n \mu_{B_i}(y) \right] dy}{H(1, n, \wedge, \vee)} \end{aligned}$$

Therefore there should be a probability space $(Y, \mathcal{F}, P)$ and a random variable η defined on

$(Y, \mathcal{F}, P)$, which obeys the probability density function $g(y)$, where $\mathcal{F}$ is a Borel σ —field on Y . If we put $\Omega \triangleq \{x_0\} \times Y$, then $\mathcal{F}$ can be regarded as a Borel σ —field on η as a random variable defined on Ω , and P as a probability on $(\Omega, \mathcal{F}, P)$. Thus we get a stochastic system

$$\underline{g} = S(X, Y, g(x_0, y)),$$

that its input–output relation is also

$$\underline{g}(x) = \int_Y yp(x_0, y)dy / \int_Y p(x_0, y)dy.$$

Let $\Delta y_j \triangleq y_j - y_{j-1} (j = 1, 2, \dots, n)$. We have

$$\begin{aligned} \underline{g}(x_0) &= \frac{\int_Y p(x_0, y)dy}{\int_Y p(x_0, y)dy} = \frac{\int_a^b \left[\bigvee_{i=1}^n \mu_{B_i}(y) \right] dy}{\int_a^b \left[\bigvee_{i=1}^n \mu_{B_i}(y) \right] dy} \\ &\approx \frac{\sum_{j=1}^n \left[\bigvee_{i=1}^n \mu_{B_i}(y_j) \right] y_j \Delta y_j}{\sum_{j=1}^n \left[\bigvee_{i=1}^n \mu_{B_i}(y_j) \right] \Delta y_j} = \frac{\sum_{j=1}^n y_j \Delta y_j}{b-a} = \sum_{j=1}^n a_j y_j, \\ a_j &\triangleq \Delta y_j / (b-a), \quad j = 1, 2, \dots, n \end{aligned} \quad (42)$$

So $\underline{g}(x_0)$ corresponding to x_0 is approximately equal to the weighted average of $y_1, y_2, \dots, y_n$. Especially, when the partition $a = y_0 < y_1 < \dots < y_n = b$ is equidistant, i.e.,

$$\Delta y_j = h, j = 1, \dots, n,$$

we also have the following expression:

$$\begin{aligned} \underline{g}(x_0) &= \frac{\int_Y p(x_0, y)dy}{\int_Y p(x_0, y)dy} \\ &\approx \frac{\sum_{j=1}^n \left[\bigvee_{i=1}^n \mu_{B_i}(y_j) \right] y_j \Delta y_j}{\sum_{j=1}^n \left[\bigvee_{i=1}^n \mu_{B_i}(y_j) \right] \Delta y_j} = \frac{\sum_{j=1}^n y_j}{n}, \end{aligned} \quad (43)$$

i.e., $\underline{g}(x_0)$ corresponding to x_0 is approximately equal to the arithmetic average of $y_1, y_2, \dots, y_n$.

Example 4 Consider a shooting practice system. For simplification, suppose that there is only one gun in the system. Every experiment, i.e., every shooting, only one bullet is shot to a target and the bullet is denoted by x_0 . Because the same kind of guns shoot the same kind of bullets and the bullets in the same kind of bullets are all regarded as x_0 , we get the input universe of the system as $X \triangleq \{x_0\}$. Every shooting that the bullet x_0 is shot to the target is regarded as putting an input to the system, and the point of impact in the target is thought as the response of the system to the input x_0 . How to measure the system response? There are many methods. Here we take the distance between the point of impact y and the center of the target as the output of the system. If we ignore missing the target of shooting, the distance between the point of impact and the center of the target is surely bounded. A felicitous upper bound is denoted by b , for example, b may be the distance between the center of the target and the edge of the target. The lower bound of the distance between the point of impact and the center of the target is clearly zero, denoted by $a = 0$. And we get an output universe $Y = [a, b]$. Since we do not in advance know which y_0 in Y should correspond to x_0 when x_0 is put into the system as an input, this is a pure stochastic system. Suppose we test the shooting level of a shooting team with n shooters. By some shooting practices, we can find the distances between the point of impact and the center of the target of the n shooters are respectively as about y_1 , about $y_2, \dots$, about y_n . We can assume that

$$a = y_0 < y_1 < \dots < y_n = b.$$

After $y_0, y_1, \dots, y_n$ are fuzzified to get fuzzy sets as follows:

$$B_i \in \mathcal{F}(Y), i = 0, 1, \dots, n,$$

which are demanded to be a fuzzy partition of Y ,

$$\mathcal{B} \triangleq \{B_i | i = 1, 2, \dots, n\}$$

is formed. Then we let

$$\begin{aligned} A_i &\triangleq X = \{x_0\}, \\ \mu_{A_i}(x) &= \chi_X(x) = \chi_{\{x_0\}}(x), \\ i &= 1, 2, \dots, n, \end{aligned}$$

and $\mathcal{A} \triangleq \{A_i | i = 1, 2, \dots, n\}$. So a fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$ is got. And we obtain a fuzzy system as being $\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$. This comes back to the way of obtaining (41), and the rest is clear.

We turn to discuss problems of fuzzy reasoning representations with one-dimensional random variables and their approximations to the stochastic systems. Also two typical cases are considered.

Typical Case 1*

This case is the contraposition of the above typical case 1. Given a stochastic system

$$\underline{s} = (X, Y \triangleq \{y_0\}, f(x) \triangleq f(x, y_0)),$$

this means there are a probability space $(\Omega, \mathcal{F}, P)$ and a random variable ξ defined on $(\Omega, \mathcal{F}, P)$ obeys probability density function $f(x) \triangleq f(x, y_0)$, where the probability space is $\Omega \triangleq X \times \{y_0\}$. From the above discussion, we know that the input–output relation of the system is as follows:

$$\begin{aligned} \underline{s}(x) &= \frac{\int_Y f(x, y_0) y_0 dy}{\int_Y f(x, y_0) dy} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\int_{y_0}^{y_0+\varepsilon} f(x, y_0) y dy}{\int_{y_0}^{y_0+\varepsilon} f(x, y_0) dy} = y_0, x \in X \end{aligned}$$

where $(\forall x \in X)(f(x, y_0) > 0)$ is supposed, and so

$$(\forall x \in X) \left(\int_{y_0}^{y_0+\varepsilon} f(x, y_0) dy > 0 \right).$$

Make a partition of X : $a = x_1 < x_2 < \dots < x_n = b$ and the nodes $x_i (i = 1, 2, \dots, n)$ are fuzzified to get fuzzy sets $A_i \in \mathcal{F}(X) (i = 1, 2, \dots, n)$, which are demanded to be a fuzzy partition of X . And let

$$\begin{aligned} B_i &\triangleq Y = \{y_0\}, \\ \mu_{B_i}(y) &= \chi_Y(y) = \chi_{\{y_0\}}(y), \\ i &= 1, 2, \dots, n, \end{aligned}$$

$\mathcal{A} \triangleq \{A_i | i = 1, 2, \dots, n\}$ and $\mathcal{B} \triangleq \{B_i | i = 1, 2, \dots, n\}$. We obtain a fuzzy inference rule group: $\mathcal{A} \rightarrow \mathcal{B}$. Thus a fuzzy system $\bar{s} = (X, Y = \{y_0\}, \mathcal{A}, \mathcal{B})$ is got. Because $\mathcal{A} = \{A_i | i = 1, 2, \dots, n\}$ is a fuzzy partition of X , and $(\forall x \in X) \left(p(x, y_0) = \sum_{i=1}^n \mu_{A_i}(x) > 0 \right)$, so

$$(\forall x \in X) \left(\int_{y_0}^{y_0+\varepsilon} p(x, y_0) dy > 0 \right).$$

The input–output relation of the fuzzy system is as follows:

$$\begin{aligned} \bar{s}(x) &= \frac{\int_Y p(x, y_0) y_0 dy}{\int_Y p(x, y_0) dy} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\int_{y_0}^{y_0+\varepsilon} p(x, y_0) y dy}{\int_{y_0}^{y_0+\varepsilon} p(x, y_0) dy} = y_0 \end{aligned}$$

Clearly $(\forall x \in X)(\bar{s}(x) = \underline{s}(x))$, which means that the fuzzy system $\bar{s}$ can accurately approximate the stochastic system $\underline{s}$.

Typical Case 2*

This case is the contraposition of the above typical case 2. Given a stochastic system:

$$\underline{s} = (X \triangleq \{x_0\}, Y \triangleq [a, b], g(y) \triangleq g(x_0, y)),$$

this means there are a probability space $(\Omega, \mathcal{F}, P)$ and a random variable η defined on $(\Omega, \mathcal{F}, P)$ obeys probability density function $g(y) = g(x_0, y)$, where the probability space is $\Omega \triangleq \{x_0\} \times Y$. Noticing that

$$\int_Y g(x_0, y) dy = \int_Y g(y) dy = 1,$$

from the above discussion, we know that the input–output relation of the system is as follows:

$$\begin{aligned}\underline{s}(x_0) &= \frac{\int_Y g(x_0, y) y dy}{\int_Y g(x_0, y) dy} \\ &= \int_a^b g(x_0, y) y dy = \int_a^b g(y) y dy, \\ &= E(\eta)\end{aligned}$$

where the output of the system $\underline{s}(x_0)$ is just the mathematical expectation of random variable η : $E(\eta)$.

Theorem 5 Given a continuous stochastic system:

$$\underline{s} = (X = \{x_0\}, Y = [a, b], g(y) = g(x_0, y)),$$

there must exist a group of fuzzy inference rules $\mathcal{A} \rightarrow \mathcal{B}$, where

$$\begin{aligned}\mathcal{A} &= \{A_i | i = 1, 2, \dots, n\}, \\ \mathcal{B} &= \{B_i | i = 1, 2, \dots, n\}, \\ A_i &\in \mathcal{F}(X), B_i \in \mathcal{F}(Y), \\ i &= 1, 2, \dots, n,\end{aligned}$$

such that the fuzzy system $\bar{s}$ constructed by the group of fuzzy inference rules can approximate the continuous stochastic system $\underline{s}$ to arbitrarily given precision.

Proof First we make a partition of Y :

$$a = y_0 < y_1 < \dots < y_n = b$$

and triangle fuzzy sets

$$B_i^* \in \mathcal{F}(Y), i = 0, 1, \dots, n.$$

Let $M \triangleq \max \{g(y) | y \in Y\}$. Then these y_i are fuzzified to become fuzzy sets:

$$\mu_{B_i}(y) \triangleq \frac{\mu_{B_i^*}(y) g(y)}{M}, \quad i = 0, 1, \dots, n.$$

Clearly $B_i \in \mathcal{F}(Y) (i = 0, 1, \dots, n)$. Put

$$A_i \triangleq \{x_0\}, i = 1, 2, \dots, n,$$

and take

$$\mathcal{A} \triangleq \{A_i | i = 1, 2, \dots, n\},$$

$$\mathcal{B} \triangleq \{B_i | i = 1, 2, \dots, n\}$$

We have a fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$. Then a fuzzy system as follows

$$\bar{s} = (X = \{x_0\}, Y = [a, b], \mathcal{A}, \mathcal{B})$$

is got, the input-output relation of which is as follows:

$$\begin{aligned}\bar{s}(x_0) &= \frac{\int_a^b p(x_0, y) y dy}{\int_a^b p(x_0, y) dy} \\ &= \frac{\int_a^b \left(\bigvee_{i=1}^n (\mu_{A_i}(x_0) \wedge \mu_{B_i}(y)) \right) y dy}{\int_a^b \left(\bigvee_{i=1}^n (\mu_{A_i}(x_0) \wedge \mu_{B_i}(y)) \right) dy} \\ &= \frac{\int_a^b \left(\bigvee_{i=1}^n \mu_{B_i}(y) \right) y dy}{\int_a^b \left(\bigvee_{i=1}^n \mu_{B_i}(y) \right) dy} = \frac{\int_a^b \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y) g(y)}{M} \right) y dy}{\int_a^b \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y) g(y)}{M} \right) dy} \\ &\approx \frac{\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) y_j \Delta y_j}{\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) \Delta y_j} \\ &= \frac{\sum_{j=1}^n \mu_{B_i^*}(y_j) g(y_j) y_j \Delta y_j}{\sum_{j=1}^n \mu_{B_i^*}(y_j) g(y_j) \Delta y_j} \\ &= \frac{\sum_{j=1}^n g(y_j) y_j \Delta y_j}{\sum_{j=1}^n g(y_j) \Delta y_j} \approx \frac{\int_a^b g(y) y dy}{\int_a^b g(y) dy} = E(\eta) = \underline{s}(x_0).\end{aligned}$$

Because $\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) y_j \Delta y_j$ and the following expression:

$$\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) \Delta y_j$$

are, respectively, Riemann sums of $\int_a^b p(x_0, y)dy$ and $\int_a^b p(x_0, y)dy$, and the following two sums

$$\sum_{j=1}^n g(y_j) y_j \Delta y_j, \quad \sum_{j=1}^n g(y_j) \Delta y_j$$

are, respectively, Riemann sums of the integrals $\int_a^b g(y)dy$ and $\int_a^b g(y)dy$, based on lemma 3, for any given an approximation precision $\varepsilon > 0$, there exists $\delta > 0$; when

$$\lambda = \max \{ \Delta y_i = y_i - y_{i-1} | i = 1, 2, \dots, n \} < \delta,$$

we simultaneously have the following inequality:

$$\left| \frac{\int_a^b p(x_0, y)dy}{\int_a^b p(x_0, y)dy} - \frac{\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) y_j \Delta y_j}{\sum_{j=1}^n \left(\bigvee_{i=1}^n \frac{\mu_{B_i^*}(y_j) g(y_j)}{M} \right) \Delta y_j} \right| < \frac{\varepsilon}{2},$$

$$\left| \frac{\int_a^b g(y)dy}{\int_a^b g(y)dy} - \frac{\sum_{j=1}^n g(y_j) y_j \Delta y_j}{\sum_{j=1}^n g(y_j) \Delta y_j} \right| < \frac{\varepsilon}{2}$$

From these, it is easy to know the result:

$$|\bar{s}(x_0) - \underline{s}(x_0)| < \varepsilon.$$

Q.E.D.

Unification on Uncertainty Systems

In the entry, an important fact is revealed that for arbitrarily given a fuzzy system $\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$ there is a probability space $(\Omega, \mathcal{F}, P)$ and a random vector (ξ, η) defined on $(\Omega, \mathcal{F}, P)$ such that (ξ, η) obeys the probability density $f(x, y)$. Thus we can get a stochastic system $\underline{s} = S(X, Y, f(x, y))$. Then based on the conclusions of this entry, for arbitrarily given a stochastic system $\underline{s}$, we can always obtain a group of fuzzy inference rules $\mathcal{A} \rightarrow \mathcal{B}$ such that a fuzzy system:

$$\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$$

can be formed by using $\mathcal{A} \rightarrow \mathcal{B}$. And the above section tells us that there is reducibility in the transformations between fuzzy systems and stochastic systems. This means that fuzzy systems and stochastic systems are unified under the system viewpoint. They look like two weights with same weight in the trays of a balance, one on a tray looking older and another one on other tray looking newer. Here older one means probability theory, and newer one is just fuzzy system theory. They have their special merits and support each other.

It is worthy of indicating that, for an uncertainty system, if you want to use probability theory to solve the problem, it is very difficult to get a kind of probability distribution, but you can try to use fuzzy system method to overcome it because obtaining a group of fuzzy inference rules is not so difficult, and sometimes it is easy. When you get a group of fuzzy inference rules, you can immediately transform the group of fuzzy inference rules into a probability density. So this is a very interesting thing.

Conclusions

The present entry has discussed a kind of united theory of uncertainty systems. As we know, the theories or methods dealing with uncertainty systems are usually of using probability theory and fuzzy set theory. It is interesting to communicate the relationship between probability theory and fuzzy set theory with respect to uncertainty systems. Firstly, we studied the probability representation problem of fuzzy systems in detail and indicated that there exists close relation between fuzzy systems and probability theory. Secondly, the fuzzy reasoning significance of stochastic systems is revealed. The main results are as follows:

1. For arbitrarily given a stochastic system as follows

$$\underline{s} = S(X, Y, f(x, y)),$$

we can always obtain a group of fuzzy inference rules $\mathcal{A} \rightarrow \mathcal{B}$ such that a fuzzy system as follows

$$\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B})$$

can be formed by using $\mathcal{A} \rightarrow \mathcal{B}$ and $\bar{s}$ can approximate $\underline{s}$ to arbitrarily given precision.

2. It is pointed out that there is reducibility in the transformations between fuzzy systems and stochastic systems. On one side, given a fuzzy system

$$\bar{s} = S(X, Y, \mathcal{A}, \mathcal{B}),$$

the reducibility is complete. On the other side, under knowing a stochastic system $\underline{s} = S(X, Y, f(x, y))$, the reducibility is approximate.

3. On one side, we have shown that for arbitrarily given a fuzzy system $\bar{s}$, its fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$ can be transformed into a probability density f of a stochastic system $\underline{s}$. On the other side, the case is just opposite, i.e., for arbitrarily given a stochastic system $\underline{s}$, its probability density f can be transformed into a fuzzy inference rule group $\mathcal{A} \rightarrow \mathcal{B}$ of a fuzzy system $\bar{s}$. Therefore, with respect to an uncertainty system, the probability density of the stochastic system and fuzzy inference rule group of the fuzzy system can be transformed into each other. In other words, with respect to an uncertainty system, fuzziness and randomness are two different sides and regarded as two different description approaches, and they essentially belong to the same thing: uncertainty. We may regard fuzziness and randomness as interlacing together, and it is hard to separate them into two self-governed parts. They look like two weights with same weight in the trays of a balance, one on a tray looking older and another one on other tray looking newer. Here older one means probability theory, and newer one is just fuzzy system theory. They have their special merits and support each other but no exclude.
4. For an uncertainty system, if you want to use probability theory to solve the problem, it is very difficult to get a kind of probability distribution, but you can try to use fuzzy system

method to overcome it because obtaining a group of fuzzy inference rules is not so difficult, and sometimes it is easy. When you get a group of fuzzy inference rules, you can immediately transform the group of fuzzy inference rules into a probability density. In this way, a lot of good tools in probability theory can be used to treat the uncertainty system. So this is a very interesting thing.

5. Just with the COG method, the relation between fuzzy systems and probability theory has been established. From the viewpoint of methodology, in a certain bound, one may use the method of probability theory to investigate fuzzy systems. From the viewpoint of philosophy, uncertainty originally contains randomness as well as fuzziness. Randomness and fuzziness are often interwoven, so it is very difficult to separate them.

Bibliography

- Dubois D, Prade H (1991a) Fuzzy sets in approximate reasoning, Part 1: Inference with possibility distributions. *Fuzzy Sets Syst* 40(1):143–202
- Dubois D, Prade H (1991b) Fuzzy sets in approximate reasoning, Part 2: logical approaches. *Fuzzy Sets Syst* 40(1):203–244
- Guo FF, Chen TY, Xia ZQ (2003) Triple I methods for fuzzy reasoning based on maximum fuzzy entropy principle. *Fuzzy Syst Math (in Chinese)* 17(4):55–59
- Hou J, You F, Li HX (2005) Some fuzzy controllers constructed by triple I method and their response capability. *Prog Natl Sci (in Chinese)* 15(1):29–37
- Li HX (1995) To see the success of fuzzy logic from mathematical essence of fuzzy control. *Fuzzy Syst Math (in Chinese)* 9(4):1–14
- Li HX (1998) Interpolation mechanism of fuzzy control. *Sci China (Series E)* 41(3):312–320
- Li HX, You F, Peng JY (2004) Fuzzy controllers based on some fuzzy implication operators and their response functions. *Prog Nat Sci* 14(1):15–20
- Pei DW (2001) Two triple I methods for FMT problem and their reductivity. *Fuzzy Syst Math (in Chinese)* 15(4):1–7
- Song SJ, Wu C (2002a) Reverse triple I method of fuzzy reasoning. *Sci China (Series F)* 45(5):344–364
- Song SJ, Wu C (2002b) Reverse triple I method with restrictions of fuzzy reasoning. *Prog Natl Sci (in Chinese)* 12(1):95–100
- Song SJ, Feng CB, Wu CX (2001) Theory of restriction degree of triple I method with total inference rules of fuzzy reasoning. *Prog Nat Sci* 11(1):58–66

- Wang GJ (1997) A formal deductive system of fuzzy propositional calculus. *Chinese Sci Bull* (in Chinese) 42(10):1041–1045
- Wang GJ (1999a) A new method for fuzzy reasoning. *Fuzzy Syst Math* (in Chinese) 13(3):1–10
- Wang GJ (1999b) Full implication triple I method for fuzzy reasoning. *Sci China (Series E)* (in Chinese) 29(1):43–53
- Wang GJ (2000) Non-classical mathematical logic and approximate reasoning (in Chinese). Science Press, Beijing
- Wang PZ, Li HX (1995) Fuzzy systems theory and Fuzzy computers (in Chinese). Science Press, Beijing
- Wang GJ, Song QY (2003) A new kind of triple I method and its logical foundation. *Prog Natl Sci* (in Chinese) 13(6):575–581
- Wu WM (1994) Principle and methods of Fuzzy reasoning (in Chinese). Guizhou Science and Technology Press, Guiyang
- You F, Li HX (2004) Fuzzy implication operators and their construction (IV). *J Beijing Normal Univ* (in Chinese) 40(5):588–599
- You F, Feng YB, Li HX (2003) Fuzzy implication operators and their construction (I). *J Beijing Normal Univ* (in Chinese) 39(5):606–611
- You F, Feng YB, Wang JY, Li HX (2004a) Fuzzy implication operators and their construction (II). *J Beijing Normal Univ* (in Chinese) 40(2):168–176
- You F, Yang XY, Li HX (2004b) Fuzzy implication operators and their construction (III). *J Beijing Normal Univ* (in Chinese) 40(4):427–432
- Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans Syst Man Cybernet* 3:28–44
- Zadeh LA (1974a) The concept of a linguistic variable and its applications to approximate reasoning (I). *Inf Sci* 8(2):199–249
- Zadeh LA (1974b) The concept of a linguistic variable and its applications to approximate reasoning (II). *Inf Sci* 8(3):301–357
- Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning (III). *Inf Sci* 9(1):43–80
- Zhang WX, Liang GX (1998) Fuzzy control and systems (in Chinese). Xi'an Jiaotong University Press, Xi'an



Function Approximation Property of Fuzzy Systems and Its Error Analysis

Hong-Xing Li

School of Control Science and Engineering,
Dalian University of Technology, Dalian,
P. R. China

Article Outline

Glossary

Definition of the Subject

Introduction

Structures of Fuzzy Systems

Function Approximation Property of Fuzzy
Systems

The Approximate Remainder Formula of $f_n(x)$ to
Continuous Function $s(X)$

The Error Estimation Between Fuzzy System
 $\bar{f}_n(x)$ and $f_n(x)$

Simulation Example

Conclusions

Bibliography

Glossary

Fuzzy systems The systems modeled by means of a group of fuzzy inference rules that are formed by using fuzzy sets.

Function approximation A complicated function in an infinite dimension function space is approximated by using a group of simple functions in a finite dimension subspace of the infinite dimension function space.

Definition of the Subject

Fuzzy systems are of a kind of ability for function approximation. When we use a fuzzy system to approximate a function, there must be some error of this approximation. Thus we need to analyze

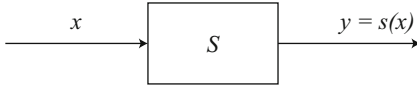
and evaluate this kind of error so that we can design a better fuzzy system to finish this approximation task.

Introduction

Since fuzzy sets were proposed by Zadeh, fuzzy systems theory has been successfully used in many fields. In practical applications, based on a given approximation accuracy we can usually establish a fuzzy system to approximate a predetermined model or control process. Therefore, the research of function approximation property of fuzzy systems has become an important direction. The so-called function approximation of fuzzy systems means that we consider whether the fuzzy system can approximate any continuous function on a compact set in any degree of accuracy (by some sort of norm). From the mathematical view, a fuzzy system is regarded as a mapping from the input universe to the output universe; especially it is an interpolator. Literature (Li 2006) reveals the probability meaning of center of gravity method, and shows that the center of gravity method is reasonable. The function approximation properties of fuzzy systems constructed by the center of gravity defuzzification method are focused on in this entry, where the approximation error and the remainder expression are regarded as very important and interesting. At last, the remainder expressions of error upper bound estimate are proved. And a simulation test is also given.

Structures of Fuzzy Systems

We first consider a static open-loop system with one input one output, i.e., SISO, as shown in Fig. 1. The value of input variable x takes its values from input universe X , and the value of output variable y takes its values from output universe Y . If the system S is a deterministic system, we can use conventional methods to establish the mathematical model of the system (for example, using method



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 1 SISO static open-loop system

of mechanism to establish differential equation model), and then the solution of the model is obtained by using analytical methods or numerical methods. So we have basically mastered the system (depthier problem is the qualitative problem of the system: controllability, observability, stability, etc.). Then, the system can be simply expressed as a function, denoted by S , that is

$$s : X \rightarrow Y, x \mapsto y \triangleq s(x) \quad (1)$$

However, facing an uncertain system, we cannot use conventional methods to establish the “exact” mathematical model of the system (usually mean the differential equation model), so it’s difficult to obtain a function of (1).

For the uncertain system, we usually do some tests to gain a group of input–output data of the system, denoted by

$$\text{IOD} \triangleq \{(x_i, y_i) \in X \times Y | i = 0, 1, \dots, n\},$$

which is called as the based data sets of the system. By means of IOD we can get a function:

$$s_1 : X^0 \rightarrow Y^0, x_i \mapsto s_1(x_i) = y_i, \\ i = 0, 1, \dots, n$$

where $X^0 = \{x_0, x_1, \dots, x_n\}$ and also we can write that $Y^0 = \{y_0, y_1, \dots, y_n\}$. Clearly, the data sets IOD of the system is exactly the graph of function s_1 .

In general, we cannot obtain the accurate mapping as being $s : X \rightarrow Y, x \mapsto y = s(x)$ but only through the data sets IOD. However, we can use the data set IOD to construct a function to indicate the input output relation of the system, as the following:

$$\bar{s}_n : X \rightarrow Y, x \mapsto y \triangleq \bar{s}_n(x) \quad (2)$$

We try to make the mapping $\bar{s}_n : X \rightarrow Y$ being close to our goal function $s : X \rightarrow Y$ such that it

satisfies the condition $\|s - \bar{s}_n\| < \varepsilon$, where $\varepsilon > 0$ is a function approximation error that is able to meet our requirements, and $\|\cdot\|$ is a kind of norm defined in the continuous function space $C(X)$, usually the norm defined as the following:

$$(\forall s \in C(X)) (\|s\| \triangleq \|s\|_\infty \triangleq \max \{|s(x)| | x \in X\}).$$

Let $X = [a, b] \subset \mathbb{R}$ be the input universe, where $\mathbb{R}$ is real number field and $Y = [c, d] \subset \mathbb{R}$ be the output universe; without loss of generality, we can assume that the input–output data set IOD meets the following condition:

$$a = x_0 < x_1 < \dots < x_n = b,$$

$$c = y_{k_0} \leq y_{k_1} \leq \dots \leq y_{k_n} = d$$

where $k_i = \sigma(i)$, and σ is a substitution, i.e., a bijection as follows

$$\sigma : \{0, 1, \dots, n\} \rightarrow \{0, 1, \dots, n\}, i \mapsto \sigma(i) \\ = k_i, \quad (3)$$

where

$$\sigma = \begin{pmatrix} 0 & 1 & \dots & n \\ k_0 & k_1 & \dots & k_n \end{pmatrix}$$

If it is without substitution, the output data set as follows

$$Y^0 = \{y_i | i = 0, 1, \dots, n\}$$

does not meet strict order relationship:

$$c = y_1 < y_2 < \dots < y_n = d$$

This strict order relationship is important in the construction of fuzzy sets $B_i \in \mathcal{F}(Y)$ ($i = 0, 1, \dots, n$).

We can also get a group of fuzzy sets as follows:

$$\mathcal{A} \triangleq \{A_i | i = 0, 1, \dots, n\}$$

by using input data set X^0 , and similarly, can get a group of fuzzy sets $\mathcal{B} \triangleq \{B_{k_i} | i = 0, 1, \dots, n\}$ by

using output data set Y^0 . These fuzzy sets A_i, B_i can be defined as a kind of triangular wave functions. Noticing that $\sigma\sigma^{-1}(i) = i$, so $k_{\sigma^{-1}(i)} = i$, and then $B_i = B_{k_{\sigma^{-1}(i)}}$, as a result, B_i is easily defined, that is $\mathcal{B} = \{B_i | i = 0, 1, \dots, n\}$.

According to the data set IOD, we can get a group of fuzzy inference rules of the system:

$$\begin{aligned} &\text{If } x \text{ is } A_i \text{ then } y \text{ is } B_i, \\ &i = 0, 1, \dots, n \end{aligned} \quad (4)$$

where A_i and B_i are, respectively, the fuzzy sets defined on the universes X and Y , i.e.,

$$A_i \in \mathcal{F}(X), B_i \in \mathcal{F}(Y), i = 0, 1, \dots, n.$$

Every inference rule can determine a fuzzy relation $R_i \in \mathcal{F}(X \times Y)$ as the following:

$$\mu_{R_i}(x, y) \triangleq \theta(\mu_{A_i}(x), \mu_{B_i}(y)), (x, y) \in X \times Y,$$

where $\theta : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a fuzzy implication operator. Usually θ can be taken as the following forms:

$\theta = \wedge$ or $\theta = \cdot$, i.e.,

$$\begin{aligned} &\wedge : [0, 1] \times [0, 1] \rightarrow [0, 1] \\ &(a, b) \mapsto \wedge(a, b) = a \wedge b, \\ &\cdot : [0, 1] \times [0, 1] \rightarrow [0, 1] \\ &(a, b) \mapsto \cdot(a, b) = a \cdot b \end{aligned}$$

We all know that fuzzy implication operator $\theta = \cdot$ is called Larsen implication operator. In this entry we often use Larsen implication operator. Now the group of inference rules (4) can be regarded as a function as follows

$$\begin{aligned} &s^* : \mathcal{A} \rightarrow \mathcal{B}, A_i \mapsto s^*(A_i) \triangleq B_i, \\ &i = 0, 1, \dots, n \end{aligned}$$

Based on these fuzzy relations $R_i, i = 0, 1, \dots, n$, we can get a whole fuzzy relations $R \in \mathcal{F}(X \times Y)$, where

$$\begin{aligned} R &= \bigcup_{i=0}^n R_i, \mu_R(x, y) = \bigvee_{i=0}^n \mu_{R_i}(x, y), \\ (x, y) &= X \times Y \end{aligned}$$

By using R , we can make a function as the following:

$$\begin{aligned} &s^{**} : \mathcal{F}(X) \rightarrow \mathcal{F}(Y) \\ &A \mapsto B = s^{**}(A) \triangleq A \circ R, \\ &\mu_B(y) = \bigvee_{x \in X} (\mu_A(x) \wedge \mu_R(x, y)) \\ &= \bigvee_{x \in X} \left[\mu_A(x) \wedge \left(\bigvee_{i=0}^n \mu_{R_i}(x, y) \right) \right] \end{aligned}$$

Then, in order to obtain the following mapping

$$\bar{s}_n : X \rightarrow Y, x \mapsto y = \bar{s}_n(x),$$

we can do it by two steps from the “set-set” mapping:

$$s^{**} : \mathcal{F}(X) \rightarrow \mathcal{F}(Y)$$

to obtain the “point-point” mapping $\bar{s}_n : X \rightarrow Y$.

Step 1 Transform “set-set” mapping into a “point-set” mapping:

$$\begin{aligned} &s_1 : X \rightarrow \mathcal{F}(Y) \\ &x \mapsto s_1(x) \triangleq s^{**}(\{x\}) \end{aligned} \quad (5)$$

where its membership function form is as the following: for any $(x, y) \in X \times Y$,

$$\begin{aligned} &\mu_{s_1(x)}(y) = \mu_{s^{**}(\{x\})}(y) \\ &= \bigvee_{\xi \in X} \left[\mu_{\{x\}}(\xi) \cdot \mu_R(\xi, y) \right] \\ &= \mu_R(x, y) = \bigvee_{i=0}^n (\mu_{A_i}(x) \cdot \mu_{B_i}(y)) \end{aligned} \quad (6)$$

>Because $(\forall x \in X)(s_1(x) \in \mathcal{F}(Y))$, so we should write that $B(\xi = x) \triangleq s_1(x)$, where ξ is regarded as a random variable taking its value in X ; so for any $(x, y) \in X \times Y, B(\xi = x)$ is shown as the following:

$$\begin{aligned} &\mu_{B(\xi=x)}(y) = \mu_{s_1(x)}(y) \\ &= \bigvee_{i=0}^n (\mu_{A_i}(x) \cdot \mu_{B_i}(y)) \end{aligned} \quad (7)$$

Step 2 Turn the fuzzy set $B(\xi = x) = s_1(x) \in \mathcal{F}(Y)$ into a point $y = (y(\xi))_{\xi=x} = y(x)$ in Y which is corresponding to $\xi = x$.

Literature (Li 2006) has proved that the physical center of gravity of plane rigid body is reasonable and is an optimal method in the sense of least squares. Suppose that

$$\int_Y |y \mu_{B(\xi=x)}(y)| dy < \infty, \\ 0 < \int_Y \mu_{B(\xi=x)}(y) dy < \infty$$

By using the gravity method, we have

$$y = (y(\xi))_{\xi=x} = y(x) = \frac{\int_Y y \mu_{B(\xi=x)}(y) dy}{\int_Y \mu_{B(\xi=x)}(y) dy} \quad (8)$$

This means that we have got the mapping as the following:

$$\bar{s}_n : X \rightarrow Y \\ x \mapsto s_n(x) \triangleq y(x) = \frac{\int_Y y \mu_{B(\xi=x)}(y) dy}{\int_Y \mu_{B(\xi=x)}(y) dy} \quad (9)$$

And then substituting (7) into the Eq. (9), we get

$$\bar{s}_n(x) = \frac{\int_Y y \mu_{B(\xi=x)}(y) dy}{\int_Y \mu_{B(\xi=x)}(y) dy} \\ = \frac{\int_Y y \left[\bigvee_{i=0}^n (\mu_{A_i}(x) \cdot \mu_{B_i}(y)) \right] dy}{\int_Y \left[\bigvee_{i=0}^n (\mu_{A_i}(x) \cdot \mu_{B_i}(y)) \right] dy} \quad (10)$$

Note 1 Because

$$(\forall (x, y) \in X \times Y) \left(\mu_{B(\xi=x)}(y) = \mu_R(x, y) \right),$$

the Eq. (10) can be written as a more general form:

$$\bar{s}_n(x) = \frac{\int_Y y \mu_R(x, y) dy}{\int_Y \mu_R(x, y) dy}, x \in X \quad (11)$$

And since $R = \bigcup_{i=0}^n R_i$, the fuzzy relation $\mathbf{R}$ is also regarded as obtaining from

$R_{k_i} (i = 0, 1, \dots, n)$ by the operation “ $\cup$.” In addition to the use of operation “ $\cup$,” many ways can be used. Therefore, (11) has a broader meaning. Q.E.D.

The function (11) $\bar{s}_n : X \rightarrow Y$ is called fuzzy system based on the operating process from Data to Formula by the above method.

From the data set

$$\text{IOD} = \{(x_i, y_i) \in X \times Y | i = 0, 1, \dots, n\},$$

we have already obtained an approximate function $\bar{s}_n$, which can approximate the input–output function $s : X \rightarrow Y$ of the system. However, in (11), the numerator and denominator are integral expressions, while in general, these two integrals have almost no analytical solution. To facilitate application purposes, we can use the definition of Riemann integral to simplify our method. In fact, let

$$\Delta y_{k_i} = y_{k_{i+1}} - y_{k_i}, i = 0, 1, \dots, n-1, \\ \Delta y_{k_n} = \frac{\sum_{i=0}^{n-1} \Delta y_{k_i}}{n} = \frac{d-c}{n},$$

Since $\Delta y_i = \Delta y_{k_{\sigma^{-1}(i)}}$, as a result, $\Delta y_i (i = 0, 1, \dots, n)$ are all defined. Then, according to the meaning of Riemann sum in the definite integration, we have

$$\bar{s}_n(x) = \frac{\int_Y y \mu_{B(\xi=x)}(y) dy}{\int_Y \mu_{B(\xi=x)}(y) dy} \\ \approx \frac{\sum_{i=0}^n \mu_{B(\xi=x)}(y_i) y_i \Delta y_i}{\sum_{i=0}^n \mu_{B(\xi=x)}(y_i) \Delta y_i} \\ = \frac{\sum_{i=0}^n \left[\bigvee_{j=0}^n (\mu_{A_j}(x) \cdot \mu_{B_j}(y_i)) \right] y_i \Delta y_i}{\sum_{i=0}^n \left[\bigvee_{j=0}^n (\mu_{A_j}(x) \cdot \mu_{B_j}(y_i)) \right] \Delta y_i} \quad (12) \\ = \frac{\sum_{i=0}^n \mu_{A_i}(x) y_i \Delta y_i}{\sum_{i=0}^n \mu_{A_i}(x) \Delta y_i} = \sum_{i=0}^n \frac{\mu_{A_i}(x) \Delta y_i}{\sum_{j=0}^n \mu_{A_j}(x) \Delta y_j} y_i \\ = \sum_{i=0}^n \mu_{A_i^*}(x) y_i$$

where we have been set

$$\mu_{A_i}^*(x) \triangleq \frac{\mu_{A_i}(x)\Delta y_i}{\sum_{j=0}^n \mu_{A_j}(x)\Delta y_j}, \quad (13)$$

$$i = 0, 1, \dots, n$$

It's not difficult to verify that the group of functions as being $\{\mu_{A_i}^*(x)\}_{i=0}^n$ is linear independent in the function space $C(X)$, and every base function $\mu_{A_i}^*(x)$ has Kronecker property:

$$\mu_{A_i}^*(x_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$i, j \in \{0, 1, \dots, n\}.$$

Let

$$f_n(x) \triangleq \sum_{i=0}^n \mu_{A_i}^*(x)y_i \quad (14)$$

Then the function $f_n(x) \triangleq \sum_{i=0}^n \mu_{A_i}^*(x)y_i$ happens to be an interpolation function, which the group of functions as being $\{\mu_{A_i}^*(x)\}_{i=0}^n$ is just regarded its base functions.

In particular, when $\Delta y_i = h$ ($i = 0, 1, \dots, n$), i.e.,

$$Y^0 = \{y_i | i = 0, 1, \dots, n\}$$

is an equidistant partition data set, where its common interval is $h > 0$, and it's not difficult to verify that $\sum_{i=0}^n \mu_{A_i}(x) \equiv 1$ then we have the following equation:

$$\begin{aligned} \mu_{A_i}^*(x) &= \frac{\mu_{A_i}(x)\Delta y_i}{\sum_{j=0}^n \mu_{A_j}(x)\Delta y_j} \\ &= \frac{\mu_{A_i}(x)}{\sum_{j=0}^n \mu_{A_j}(x)} = \mu_{A_i}(x), \end{aligned} \quad (15)$$

$$i = 0, 1, \dots, n$$

And then the Eqs. (12) and (14) will be simplified as

$$s(x) \approx \bar{s}_n(x) \approx f_n(x) = \sum_{i=0}^n \mu_{A_i}(x)y_i \quad (16)$$

This means that $\bar{s}_n(x)$ is approximately a piecewise linear interpolation function as the following:

$$f_n(x) = \sum_{i=0}^n \mu_{A_i}(x)y_i$$

So $s(x)$ can be approximately regarded as a piecewise linear interpolation function $f_n(x)$.

From the data set IOD, we write the following symbol:

$$\text{FIOD} \triangleq \{(A_i, B_i) \in \mathcal{F}(X) \times \mathcal{F}(Y) | i = 0, 1, \dots, n\}.$$

If it satisfies the condition:

$$(i = 0, 1, \dots, n)(\mu_{A_i}(x) \in C(X), \mu_{B_i}(y) \in C(Y)),$$

then FIOD is called continuous fuzzy data set. We call FIOD being with two-phase property, if it satisfies the condition:

$$\begin{aligned} \sum_{i=0}^n \mu_{A_i}(x) &\equiv 1, \sum_{i=0}^n \mu_{B_i}(x) \equiv 1, \\ (\forall x \in X)(\exists i \in \{0, \dots, n-1\})(\mu_{A_i}(x) + \mu_{A_{i+1}}(x) &= 1) \\ (\forall y \in X)(\exists j \in \{0, \dots, n-1\})(\mu_{B_j}(x) + \mu_{B_{j+1}}(x) &= 1) \end{aligned}$$

Clearly, the two-phase fuzzy data set satisfies the Kronecker properties:

$$A_i(x_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad B_i(y_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

where $i, j \in \{0, 1, \dots, n\}$.

Note 2 It's not difficult to verify when the FIOD is with two-phase property, the conclusions derived above remain valid. **Q.E.D.**

Function Approximation Property of Fuzzy Systems

For the data set

$$\text{IOD} = \{(x_i, y_i) \in X \times Y | i = 0, 1, \dots, n\},$$

write

$$\Delta x_i \triangleq x_{i+1} - x_i, \quad i = 0, 1, \dots, n-1,$$

obviously, $\max_{0 \leq i \leq n-1} \{\Delta x_i\} \rightarrow 0 \Rightarrow n \rightarrow \infty$; on the contrary, that $n \rightarrow \infty \Rightarrow \max_{0 \leq i \leq n-1} \{\Delta x_i\} \rightarrow 0$ is not true. And if that $n \rightarrow \infty \Rightarrow \max_{0 \leq i \leq n-1} \Delta x_i \rightarrow 0$ is true, then data set IOD is called coordinated. That is, IOD is coordinated if and only if

$$n \rightarrow \infty \Leftrightarrow \max_{0 \leq i \leq n-1} \{\Delta x_i\} \rightarrow 0.$$

We always assume that the following data set IOD is coordinated.

In addition, for a continuous function $s \in C[a, b]$, if the data set IOD meet the interpolation condition about s , i.e.,

$$(\forall i \in \{0, 1, \dots, n\})(y_i = s(x_i)),$$

then it's not difficult to understand that IOD is coordinated, which implies $n \rightarrow \infty \Leftrightarrow$

$$\max_{0 \leq i \leq n} \{\Delta y_i\} \rightarrow 0.$$

For a given data set IOD, we can get its FIOD with two-phase property as follows:

$$\text{FIOD} = \{(A_i, B_i) \in \mathcal{F}(X) \times \mathcal{F}(Y) | i = 0, 1, \dots, n\}.$$

Suppose the fuzzy system $\bar{s}_n$ like (11) to be constructed by IOD and IODF, that is, we have got the equation as the following:

$$\bar{s}_n(x) = \frac{\int_Y y \mu_R(x, y) dy}{\int_Y \mu_R(x, y) dy}, \quad x \in X \quad (17)$$

And we also suppose that it satisfies condition:

$$0 < \int_Y \mu_R(x, y) dy < \infty, \quad \int_Y |\mu_R(x, y)| dy < \infty.$$

We call the fuzzy system $\bar{s}_n$ has function approximation property in $X = [a, b]$, if for any function $s \in C[a, b]$, as long as the data set IOD satisfies interpolation condition:

$$(\forall i \in \{0, 1, \dots, n\})(y_i = s(x_i)),$$

then $\bar{s}_n$ converges s according to the norm of normed linear space, that is, for any $\varepsilon > 0$, $\exists N \in \mathbb{N}^+$, such that

$$(\forall n \in \mathbb{N}^+)(n > N \Rightarrow \|s - \bar{s}_n\|_\infty < \varepsilon) \quad (18)$$

where $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$, while $\mathbb{N}$ is the set of all natural numbers; in other words, the function sequence $\{\bar{s}_n\}_{n=1}^\infty$ uniformly converges to a continuous function s in the universe $X = [a, b]$.

Lemma 1 Let $f(x, y)$ be a binary continuous function on $X \times Y$, where $X = [a, b]$ and $Y = [c, d]$ are two real number closed intervals for the integral as follows

$$I(x) = \int_c^d f(x, y) dy$$

with parameter x , we must have this result: for any $\varepsilon > 0$, there is always $\delta > 0$ which has nothing to do with the parameter x , for any partition of Y :

$$c = y_0 < y_1 < \dots < y_n = d,$$

if $\lambda \triangleq \max_{0 \leq i \leq n-1} \{\Delta y_i\} < \delta$, then for all $x \in X$, the

Riemann sum $\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i$ of $I(x)$ must uniformly meet the following strict inequality:

$$\left| I(x) - \sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i \right| < \varepsilon \quad (19)$$

Proof For any $\varepsilon > 0$, take $\delta_k = 1/k$, $k = 1, 2, \dots$. It's proved that there must be one k , such that $\delta = \delta_k$ satisfies the conclusion of the lemma. Otherwise, for every k , exist $x_k \in X$ and a partition of Y :

$$c = y_0^{(k)} < y_1^{(k)} < \cdots < y_{n_k}^{(k)} = d,$$

and also exist $\xi_i^{(k)} \in [y_i^{(k)}, y_{i+1}^{(k)}], i = 0, 1, \dots, n-1$,

although $\lambda_k = \max_{0 \leq i \leq n_k-1} \{\Delta y_i^{(k)}\} < \delta_k$, but

$$\left| I(x_k) - \sum_{i=0}^{n_k-1} f(x_k, \xi_i^{(k)}) \Delta y_i^{(k)} \right| \geq \varepsilon.$$

Attention to that $\{x_k\}$ is bounded point sequence, there must be a convergent subsequence $\{x_{k_j}\}$ of $\{x_k\}$, such that $x_{k_j} \xrightarrow{j \rightarrow \infty} x_* \in X$.

And noticing that $\delta_{k_j} \xrightarrow{j \rightarrow \infty} 0$, we have

$$\begin{aligned} 0 < \varepsilon &\leq \lim_{j \rightarrow \infty} \left| I(x_{k_j}) - \sum_{i=0}^{n_{k_j}-1} f(x_{k_j}, \xi_i^{(k_j)}) \Delta y_i^{(k_j)} \right| \\ &= \left| I(x_*) - \int_c^d f(x_*, y) dy \right| = 0 \end{aligned}$$

This is a clear contradiction and proves the lemma. **Q.E.D.**

Lemma 2 Let $f(x, y)$ be a binary continuous function on $X \times Y$, where $X = [a, b]$ and $Y = [c, d]$ are two real number closed intervals, for the integral as follows

$$I(x) = \int_c^d f(x, y) dy$$

with parameter x , if $(\forall x \in X)(I(x) > 0)$, then there must be $\delta > 0$, such that for any partition of Y :

$$c = y_0 < y_1 < \cdots < y_n = d,$$

and any $\xi_i \in [y_i, y_{i+1}]$, the Riemann sum as follows

$$\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i$$

of $I(x)$ must satisfy the following implication:

$$\begin{aligned} \lambda &= \lim_{0 \leq i \leq n-1} \{\Delta y_i\} < \delta \Rightarrow \\ (\forall x \in X) &\left(\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i > 0 \right) \end{aligned} \quad (20)$$

Proof It is easy to know that

$$I(x) = \int_c^d f(x, y) dy \in C[a, b];$$

therefore, there must be the minimal point $x_0 \in X$ of $I(x)$, such that the following expression is true:

$$(\forall x \in X)(I(x) \geq I(x_0)).$$

Take $\varepsilon = I(x_0)$, according to Lemma 1, exist $\delta > 0$, for any partition of Y :

$$c = y_0 < y_1 < \cdots < y_n = d,$$

and any $\xi_i \in [y_i, y_{i+1}]$, the Riemann sum as follows:

$$\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i$$

of $I(x)$ must satisfy the result: if $\lambda \triangleq \max_{0 \leq i \leq n-1} \{\Delta y_i\} < \delta$, then we have

$$(\forall x \in X) \left(\left| I(x) - \sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i \right| < \varepsilon \right),$$

so that

$$\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i > I(x) - \varepsilon \geq I(x_0) - \varepsilon = 0$$

is true for all $x \in X$, therefore the conclusion of the lemma is true. **Q.E.D.**

According to Lemma 1 and 2, we can get the following Lemma 3.

Lemma 3 Let $f(x, y)$ and $g(x, y)$ be two binary continuous functions on the universe $X \times Y$, where $X = [a, b]$ and $Y = [c, d]$ are two real number closed intervals, satisfying the condition:

$$(\forall x \in X) \left(\int_c^d g(x, y) dy > 0 \right).$$

For any $\varepsilon > 0$, there is always $\delta > 0$, which has nothing to do with parameter x , for any partition

$$c = y_0 < y_1 < \cdots < y_n = d,$$

if $\lambda \triangleq \max_{0 \leq i \leq n-1} \{\Delta y_i\} < \delta$, then for all $x \in X$, we uniformly have the following inequality:

$$\left| \frac{\int_c^d f(x, y) dy}{\int_c^d g(x, y) dy} - \frac{\sum_{i=0}^{n-1} f(x, \xi_i) \Delta y_i}{\sum_{i=0}^{n-1} g(x, \xi_i) \Delta y_i} \right| < \varepsilon \quad (21)$$

where $\xi_i \in [y_i, y_{i+1}]$, $i = 0, 1, \dots, n-1$.

Proof According to Lemma 2, we know the fact that

$$\frac{1}{\sum_{i=0}^{n-1} g(x, \xi_i) \Delta y_i}$$

is with sense for the bigger $n \in \mathbb{N}^+$. According to the limit operation rules, namely, “the limit of quotient is equal to the quotient of limits,” and then by Lemma 1, we know that the conclusion of Lemma 3 is true. **Q.E.D.**

Theorem 1 With respect to the data set IOD, the fuzzy data set with two-phase property is as follows

$$\text{FIOD} = \{(A_i, B_i) \in \mathcal{F}(X) \times \mathcal{F}(Y) | i = 0, 1, \dots, n\}.$$

Suppose that the fuzzy system $\bar{s}_n$ as (11) constructed by IOD and FIOD as follows

$$\bar{s}_n(x) = \frac{\int_Y y \mu_R(x, y) dy}{\int_Y \mu_R(x, y) dy}, \quad x \in X$$

satisfies the condition:

$$0 < \int_Y R(x, y) dy < \infty \quad \int_Y |y R(x, y)| dy \leq \infty,$$

then the fuzzy system $\bar{s}_n$ has function approximation property.

Proof For any given function $s \in C[a, b]$, suppose that the data set IOD satisfies the interpolation condition:

$$(\forall i \in \{0, 1, \dots, n\}) (y_i = s(x_i)).$$

We want to prove that $\bar{s}_n$ converge to s according to the norm of normed linear space $(C[a, b], \|\cdot\|)$, that is, for any $\varepsilon > 0$, $\exists N \in \mathbb{N}^+$, such that

$$(\forall n \in \mathbb{N}^+) (n > N \Rightarrow \|s - \bar{s}_n\|_\infty < \varepsilon).$$

In fact, for any $\varepsilon > 0$, according to Eq. (14), we easily know that, $\exists N_1 \in \mathbb{N}^+$, such that

$$(\forall n \in \mathbb{N}^+) (n > N_1 \Rightarrow \|f_n - \bar{s}_n\|_\infty < \frac{\varepsilon}{2}).$$

Then, when $n > N_1$, we have that

$$\begin{aligned} \|s - \bar{s}_n\|_\infty &\leq \|s - f_n\|_\infty + \|f_n - \bar{s}_n\|_\infty \\ &< \|s - f_n\|_\infty + \frac{\varepsilon}{2}. \end{aligned}$$

So we only consider the estimation of $\|s - f_n\|_\infty$.

In fact, it's easy to test that $\left\{ \mu_{A_i^*}(x) \right\}_{i=0}^n$ satisfies the two-phase property, and for any given $x \in [a, b]$, there must be $i \in \{0, 1, \dots, n-1\}$, such that $x \in [x_i, x_{i+1}]$, then

$$f_n(x) = \mu_{A_i^*}(x) y_i + \mu_{A_{i+1}^*}(x) y_{i+1}.$$

By using $\mu_{A_i^*}(x) + \mu_{A_{i+1}^*}(x) = \sum_{i=0}^n \mu_{A_i^*}(x) \equiv 1$, we have

$$\begin{aligned}
|s(x) - f_n(x)| &= \left| s(x) - \left(\mu_{A_i^*}(x)y_i + \mu_{A_{i+1}^*}(x)y_{i+1} \right) \right| \\
&= \left| s(x) \left(\mu_{A_i^*}(x) + \mu_{A_{i+1}^*}(x) \right) - \left(\mu_{A_i^*}(x)s(x_i) + \mu_{A_{i+1}^*}(x)s(x_{i+1}) \right) \right| \\
&\leq \mu_{A_i^*}(x)|s(x) - s(x_i)| + \mu_{A_{i+1}^*}(x)|s(x) - s(x_{i+1})| \\
&\leq |s(x) - s(x_i)| + |s(x) - s(x_{i+1})| \leq 2 \max_{i \leq m \leq i+1} \{|s(x) - s(x_m)|\}.
\end{aligned}$$

Because of $s \in C[a, b]$, $s(x)$ is uniform continuous on $[a, b]$, then for the above $\varepsilon > 0$, there must be $\delta > 0$, such that for any two points $x', x'' \in [a, b]$, we have

$$|x' - x''| < \delta \Rightarrow |s(x') - s(x'')| < \varepsilon/4$$

By using the coordination of IOD, there exists a number $N_2 \in \mathbb{N}^+$, such that

$$(\forall n \in \mathbb{N}^+) \left(n > N_2 \Rightarrow \lambda = \max_{0 \leq i \leq n-1} \{\Delta x_i\} < \delta \right).$$

Noticing that $x \in [x_i, x_{i+1}]$, we can easily know that

$$\begin{aligned}
\max_{i \leq m \leq i+1} \{|x - x_m|\} &\leq \max_{i \leq m \leq i+1} \{\Delta x_m\} \\
&\leq \max_{0 \leq i \leq n-1} \{\Delta x_i\} < \delta.
\end{aligned}$$

Therefore we have

$$\begin{aligned}
|s(x) - f_n(x)| &\leq 2 \max_{i \leq m \leq i+1} \{|s(x) - s(x_m)|\} \\
&< 2 \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.
\end{aligned}$$

Because x is arbitrary in $[a, b]$, we have that

$$\max_{x \in [a, b]} \{|s(x) - f_n(x)|\} \leq \frac{\varepsilon}{2},$$

namely, $\|s - f_n\| \leq \frac{\varepsilon}{2}$. Finally, we get the fact that, there must exist the number $N = \max \{N_1, N_2\} \in \mathbb{N}^+$, such that

$$(\forall n \in \mathbb{N}^+) \left(n > N \Rightarrow \|s - \bar{s}_n\|_\infty < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \right)$$

This means that the fuzzy system $\bar{s}_n$ has function approximation property. **Q.E.D.**

The Approximate Remainder Formula of $f_n(x)$ to Continuous Function $s(x)$

Theorem 2 In the condition of Theorem 1, for any a function $s \in C^2[a, b]$, suppose that

$$(\forall i \in \{0, 1, \dots, n-1\}) (\mu_{A_i}, \mu_{A_{i+1}} \in C^2[x_i, x_{i+1}]);$$

for the function s , if the data set IOD satisfies interpolation conditions:

$$(\forall i \in \{0, 1, \dots, n\}) (y_i = s(x_i)),$$

then we have the following results:

1. The approximate remainder of $f_n(x)$ to the continuous function $s(x)$ is of the following equation:

$$\begin{aligned}
r_n(x) &= s(x) - f_n(x) \\
&= \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)} (\varphi_{i1}''(\xi_i) - \phi_{i2}''(\xi_i)), \quad (22)
\end{aligned}$$

where

$$(\forall i \in \{0, 1, \dots, n-1\}) (x \in [x_i, x_{i+1}], \xi_i \in (x_i, x_{i+1}))$$

and

$$\begin{aligned}\varphi''_{i1}(\xi_i) &= s''(\xi_i)(\mu_{A_i}(\xi_i)\Delta y_i + \mu_{A_{i+1}}(\xi_i)\Delta y_{i+1}) \\ &\quad + 2s'(\xi_i)(\mu'_{A_i}(\xi_i)\Delta y_i + \mu'_{A_{i+1}}(\xi_i)\Delta y_{i+1}) \\ &\quad + s(\xi_i)(\mu''_{A_i}(\xi_i)\Delta y_i + \mu''_{A_{i+1}}(\xi_i)\Delta y_{i+1}),\end{aligned}$$

$$\begin{aligned}\varphi''_{i2}(t) &= \mu''_{A_i}(\xi_i)s(x_i)\Delta y_i + \mu''_{A_{i+1}}(\xi_i)s(x_{i+1})\Delta y_{i+1}, \\ q_i(x) &= \mu_{A_i}(x)\Delta y_i + \mu_{A_{i+1}}(x)\Delta y_{i+1},\end{aligned}$$

2. The approximate error estimate formula of $f_n(x)$ to the continuous function $s(x)$ is as the following inequality:

$$\|r_n\|_\infty = \|s(x) - f_n(x)\|_\infty \leq M\Delta_1^3 \quad (23)$$

where $\Delta_1 \triangleq \max_{0 \leq j \leq n-1} \{\Delta x_j\}$, $\Delta_{2i} \triangleq \max_{0 \leq i \leq n-1} \{\Delta y_i, \Delta y_{i+1}\}$, and

$$C_i \triangleq \min \{q_i(x) | x \in [x_i, x_{i+1}]\},$$

$$M_i \triangleq \frac{1}{8C_i} (M_{2i} + 4M_{1i}L_{1i} + 4M_0L_{2i}),$$

$$M_0 \triangleq \max_{t \in [a, b]} \{|s(t)|\}, \quad M \triangleq \max_{0 \leq i \leq n-1} \{M_i \cdot N\}$$

$$M_{1i} \triangleq \max_{t \in [x_i, x_{i+1}]} \{|s'(t)|\}, \quad M_{2i} \triangleq \max_{t \in [x_i, x_{i+1}]} \{|s''(t)|\},$$

$$L_{1i} \triangleq \max \left\{ \max_{t \in [x_i, x_{i+1}]} \{|\mu'_{A_i}(t)|\}, \max_{t \in [x_i, x_{i+1}]} \{|\mu'_{A_{i+1}}(t)|\} \right\},$$

$$L_{2i} \triangleq \max \left\{ \max_{t \in [x_i, x_{i+1}]} \{|\mu''_{A_i}(t)|\}, \max_{t \in [x_i, x_{i+1}]} \{|\mu''_{A_{i+1}}(t)|\} \right\}.$$

Proof 1. For any a point $x \in [a, b]$, when

$$x = x_i (i = 0, 1, \dots, n),$$

the conclusion is clearly true; so we only consider the situation of the following:

$$x \neq x_i (i = 0, 1, \dots, n).$$

In fact, there must exist a number $i \in \{0, 1, \dots, n-1\}$, such that $x \in (x_i, x_{i+1})$, according to the two-phase property of A_i ($i = 0, 1, \dots, n$), we have the following equation:

$$\begin{aligned}f_n(x) &= \sum_{j=0}^n \mu_{A_j^*}(x)s(x_j) \\ &= \mu_{A_i^*}(x)s(x_i) + \mu_{A_{i+1}^*}(x)s(x_{i+1}) \\ &= \frac{\mu_{A_i}(x)s(x_i)\Delta y_i + \mu_{A_{i+1}}(x)s(x_{i+1})\Delta y_{i+1}}{\mu_{A_i}(x)\Delta y_i + \mu_{A_{i+1}}(x)\Delta y_{i+1}} \quad (24) \\ &= \frac{p_i(x)}{q_i(x)}\end{aligned}$$

of which we have set

$$p_i(x) \triangleq \mu_{A_i}(x)s(x_i)\Delta y_i + \mu_{A_{i+1}}(x)s(x_{i+1})\Delta y_{i+1} \quad (25)$$

$$q_i(x) \triangleq \mu_{A_i}(x)\Delta y_i + \mu_{A_{i+1}}(x)\Delta y_{i+1}. \quad (26)$$

As a result, $f_n(x)$ is actually a rational fraction, that is

$$f_n(x) = p_i(x)/q_i(x),$$

and thus the approximation of $f_n(x)$ to $s(x)$ is a rational fraction approximate problem. Because $f_n(x)$ makes sense, we know that $(\forall x \in X)(q_i(x) > 0)$. And then, we proceed to consider the representation of the following approximate item:

$$r_n(x) = s(x) - f_n(x) = s(x) - \frac{p_i(x)}{q_i(x)}$$

Reform it into

$$\begin{aligned}r_n(x)q_i(x) &= s(x)q_i(x) - f_n(x)q_i(x) \\ &= s(x)q_i(x) - p_i(x)\end{aligned} \quad (27)$$

According to the interpolation conditions, we know that

$$r_n(x_j) = 0, j = i, i+1,$$

or

$$s(x_j)q_i(x_j) - p_i(x_j) = 0, j = i, i + 1.$$

Assume that $r_n(x)q_i(x)$ has the form

$$r_n(x)q_i(x) = k(x)(x - x_i)(x - x_{i+1}) \quad (28)$$

That is,

$$\begin{aligned} r_n(x)q_i(x) &= k(x)(x - x_i)(x - x_{i+1}) \\ &= s(x)q_i(x) - p_i(x) \end{aligned}$$

of which function $k(x)$ is undetermined. Now we start to consider how to solve $k(x)$.

In fact, let x be a fixed point, and do an auxiliary function:

$$\begin{aligned} \varphi_i(t) &= s(t)q_i(t) - p_i(t) - k(x)(t - x_i) \\ &\quad \times (t - x_{i+1}) \end{aligned} \quad (29)$$

Obviously, $\varphi_i \in C^2[x_i, x_{i+1}]$. According to (29), we know that $\varphi_i(x_j) = 0, j = i, i + 1$; Also, it's not hard to see $\varphi_i(x) = 0$. This means that, in the closed interval $\varphi_i(t)$ has three zero points: $x_i < x < x_{i+1}$.

According to Rolle's differential intermediate value theorem, the derivate function $\varphi'_i(t)$ of $\varphi_i(t)$ in the open intervals (x_i, x) and (x, x_{i+1}) has a zero point, denoted by $\xi_{i1} = \xi_{i1}(x)$ and $\xi_{i2} = \xi_{i2}(x)$, respectively, which all rely on x .

We continue to use Rolle's differential intermediate value theorem, the derivate function $\varphi''_i(t)$ of $\varphi'_i(t)$ in the open interval (ξ_{i1}, ξ_{i2}) has a zero point, denoted by $\xi_i = \xi_i(x)$ (certainly, it also rely on x), that is $\varphi''_i(\xi_i) = 0$.

For more convenience, we write

$$\varphi_i(t) = \varphi_{i1}(t) - \varphi_{i2}(t) - \varphi_{i3}(t),$$

where

$$\begin{aligned} \varphi_{i1}(t) &\triangleq s(t)q_i(t), \quad \varphi_{i2}(t) \triangleq p_i(t), \\ \varphi_{i3}(t) &\triangleq k(x)(t - x_i)(t - x_{i+1}). \end{aligned}$$

Then we have

$$\begin{aligned} \varphi''_{i1}(t) &= s''(t)(\mu_{A_i}(t)\Delta y_i + \mu_{A_{i+1}}(t)\Delta y_{i+1}) \\ &\quad + 2s'(t)(\mu'_{A_i}(t)\Delta y_i + \mu'_{A_{i+1}}(t)\Delta y_{i+1}) \\ &\quad + s(t)(\mu''_{A_i}(t)\Delta y_i + \mu''_{A_{i+1}}(t)\Delta y_{i+1}), \end{aligned}$$

$$\varphi''_{i2}(t) = (\mu''_{A_i}(t)s(x_i)\Delta y_i + \mu''_{A_{i+1}}(t)s(x_{i+1})\Delta y_{i+1}),$$

$$\varphi''_{i3}(t) = 2k(x).$$

Now if we substitute $\varphi''_{i1}(\xi_i(x))$, $\varphi''_{i2}(\xi_i(x))$, $\varphi''_{i3}(\xi_i(x))$ into the equation $\varphi''_i(\xi_i(x)) = 0$, we have the following equation:

$$k(x) = \frac{1}{2}(\varphi''_{i1}(\xi_i(x)) - \varphi''_{i2}(\xi_i(x))).$$

And then substituting $k(x)$ into (28), we have

$$\begin{aligned} r_n(x) &= s(x) - f_n(x) = s(x) - \frac{p_i(x)}{q_i(x)} \\ &= \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)}(\varphi''_{i1}(\xi_i(x)) - \varphi''_{i2}(\xi_i(x))), \\ i &= 0, 1, \dots, n - 1. \end{aligned}$$

2. For any $x \in [x_i, x_{i+1}]$, it's easy to see that

$$\begin{aligned} &|(x - x_i)(x - x_{i+1})| \\ &\leq \left| \left(\frac{x_i + x_{i+1}}{2} - x_i \right) \left(\frac{x_i + x_{i+1}}{2} - x_{i+1} \right) \right| \leq \frac{1}{4} \Delta x_i^2. \end{aligned}$$

Because $q_i(x) > 0$ and $q_i(x)$ is continuous on closed interval $[x_i, x_{i+1}]$, there is a minimal value as the following:

$$C_i = \min \{q_i(x) | x \in [x_i, x_{i+1}]\}.$$

Let $C \triangleq \min_{1 \leq i \leq n-1} \{C_i\}$, and then, we have the following in-equation:

$$\begin{aligned} |\varphi''_{i1}(\xi_i(x))| &\leq \Delta_{2i}(M_2 + 4M_1L_1 + 2M_0L_2), \\ |\varphi''_{i2}(\xi_i(x))| &\leq 2M_0L_2\Delta_{2i}. \end{aligned}$$

Finally, according to the remainder expression, we have the following inequality:

$$\begin{aligned}
r_i &\triangleq \max_{x \in [x_i, x_{i+1}]} \{|r_n(x)|\} \\
&= \max_{x \in [x_i, x_{i+1}]} \left\{ \left| \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)} (\varphi''_{i1}(\xi_i(x)) - \varphi''_{i2}(\xi_i(x))) \right| \right\} \\
&\leq \max_{x \in [x_i, x_{i+1}]} \left\{ \left| \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)} (|\varphi''_{i1}(\xi_i(x))| + |\varphi''_{i2}(\xi_i(x))|) \right| \right\} \\
&\leq \Delta x_i^2 \frac{\Delta_{2i}}{8C_i} (M_{2i} + 4M_{1i}L_{1i} + 4M_{0i}L_{2i}).
\end{aligned}$$

Also, because $s(x)$ is continuous function, so we have the equation $\Delta y_i = s'(\xi_i)\Delta x_i$, that is, $\Delta y_i \Leftrightarrow c \cdot \Delta x_i$, where c is a constant. Therefore,

$$\|r_n\|_\infty = \max_{0 \leq i \leq n-1} \{r_i\} \leq M\Delta_1^3.$$

Q.E.D.

Corollary 1 In Theorem 2, if all $A_i (i = 0, 1, \dots, n)$ are triangular wave functions, then

1. The approximate remainder representation of $f_n(x)$ to the function $s(x)$ is shown as follows:

$$\begin{aligned}
r_n(x) &= s(x) - f_n(x) \\
&= \phi''_{i1}(\xi_i(x)) \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)} \quad (30)
\end{aligned}$$

where $i = 0, \dots, n - 1, x \in [x_i, x_{i+1}]$, $\xi_i \in (x_i, x_{i+1})$ and

$$q_i(x) = A_i(x)\Delta y_i + A_{i+1}(x)\Delta y_{i+1},$$

$$\begin{aligned}
&\varphi''_{i1}(\xi_i(x)) \\
&= s''(\xi_i(x)) \left((\xi_i(x) - x_i) \frac{\Delta y_{i+1}}{\Delta x_i} - (\xi_i(x) - x_{i+1}) \frac{\Delta y_i}{\Delta x_i} \right) \\
&\quad + 2s'(\xi_i(x)) \left(\frac{\Delta y_{i+1}}{\Delta x_i} - \frac{\Delta y_i}{\Delta x_i} \right).
\end{aligned}$$

2. The approximate error estimate expression of $f_n(x)$ to the continuous function $s(x)$ is shown as follows:

$$\|r_n\|_\infty \leq \frac{\Delta x_i \Delta_{2i}}{8C_i} (M_{2i} \Delta x_i + 4M_{1i}). \quad (31)$$

Proof Firstly, it is easy to understand the fact that, for any a number $i \in \{0, 1, \dots, n - 1\}$, we have

$$\begin{aligned}
\mu'_{A_i}(x) &= -\frac{1}{\Delta x_i}, \quad \mu'_{A_{i+1}}(x) = \frac{1}{\Delta x_i}, \\
\mu''_{A_i}(x) &= 0 = \mu''_{A_{i+1}}(x)
\end{aligned}$$

Substituting them into the expression of $\varphi''_{i1}(\xi_i(x))$ and $\varphi''_{i2}(\xi_i(x))$, we get the following inequality:

$$|\varphi''_{i1}(\xi_i(x))| \leq \Delta_{2i} \left(M_2 + 4M_{1i} \frac{1}{\Delta x_i} \right).$$

And then we have the following inequality:

$$\begin{aligned}
r_i &= \max_{x \in [x_i, x_{i+1}]} \{|r_n(x)|\} \\
&= \max_{x \in [x_i, x_{i+1}]} \left| \varphi''_{i1}(\xi_i(x)) \frac{(x - x_i)(x - x_{i+1})}{2q_i(x)} \right| \\
&\leq \frac{\Delta x_i \Delta_{2i}}{8C_i} (M_{2i} \Delta x_i + 4M_{1i}).
\end{aligned}$$

Finally, we have the inequation:

$$\|r_n\|_\infty = \max_{0 \leq i \leq n-1} r_i \leq M_s \Delta_1^2,$$

where

$$M_s \triangleq \max_{0 \leq i \leq n-1} \left\{ \frac{N}{8C_i} (M_{2i} \Delta x_i + 4M_{1i}) \right\}.$$

Q.E.D.

The Error Estimation Between Fuzzy System $\bar{s}_n(x)$ and $f_n(x)$

Now we consider the error estimation problem between fuzzy system $\bar{s}_n(x)$ and $f_n(x)$.

Theorem 3 Under the conditions of theorem 1, for any a continuous function $s \in C[a, b]$, write

$$\bar{r}_n(x) \triangleq f_n(x) - \bar{s}_n(x).$$

Suppose that

$$(\forall i \in \{0, \dots, n-1\}) (\mu_{B_i}(y), \mu_{B_{i+1}}(y) \in C^1[y_i, y_{i+1}]),$$

and $A_i, B_{i+1}, i = 0, 1, \dots, n$ are fuzzy numbers. If the data set IOD satisfies the interpolation condition:

$$(\forall i \in \{0, 1, \dots, n\}) (y_i = s(x_i)),$$

then

$$\|\bar{r}_n\|_\infty = \|f_n - \bar{s}_n\|_\infty \leq \Gamma \Delta_1^2 \quad (32)$$

where

$$\Gamma \triangleq 4 \frac{2M_0 + 2M_0 K_1 N + N}{D^2},$$

$$P_n(x) \triangleq \sum_{i=0}^n \mu_{A_i}(x) y_i \Delta y_i, \quad Q_n(x) \triangleq \sum_{i=0}^n \mu_{A_i}(x) \Delta y_i,$$

$$\bar{P} \triangleq \max_{x \in X} \{|P_n(x)|\} \leq 2M_0 \Delta_2, \quad \bar{Q} \triangleq \max_{x \in X} \{|Q_n(x)|\} \leq 2\Delta_2,$$

$$D \triangleq \min \left\{ \min_{x \in X} \{|Q_n(x)|\}, \min_{x \in X} \int_c^d \mu_R(x, y) dy \right\},$$

$$K_1 \triangleq \max_{0 \leq i \leq n-1} \left\{ \max_{y \in [y_{k_i}, y_{k_{i+1}}]} \left\{ |\mu'_{B_{k_i}}(y)| \right\}, \max_{y \in [y_{k_i}, y_{k_{i+1}}]} \left\{ |\mu'_{B_{k_{i+1}}}(y)| \right\} \right\}.$$

Proof For any a point $x \in X = [a, b]$, we have the following expression:

$$\begin{aligned} |f_n(x) - \bar{s}_n(x)| &= \left| \sum_{i=0}^n \mu_{A_{k_i}}^*(x) y_{k_i} - \frac{\int_c^d y \mu_R(x, y) dy}{\int_c^d \mu_R(x, y) dy} \right| \\ &= \left| \frac{\sum_{i=0}^n \mu_{A_{k_i}}(x) y_{k_i} \Delta y_{k_i}}{\sum_{i=0}^n \mu_{A_{k_i}}(x) \Delta y_{k_i}} - \frac{\int_c^d y \mu_R(x, y) dy}{\int_c^d \mu_R(x, y) dy} \right| = \left| \frac{P_n(x)}{Q_n(x)} - \frac{\int_c^d y \mu_R(x, y) dy}{\int_c^d \mu_R(x, y) dy} \right| \\ &= \left| \frac{Q_n(x) \left(\int_c^d y \mu_R(x, y) dy - P_n(x) \right) + P_n(x) \left(Q_n(x) - \int_c^d \mu_R(x, y) dy \right)}{Q_n(x) \int_c^d \mu_R(x, y) dy} \right| \\ &\leq \frac{|Q_n(x)| \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| + |P_n(x)| \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right|}{|Q_n(x)| \int_c^d \mu_R(x, y) dy} \\ &\leq \frac{\bar{Q} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| + \bar{P} \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right|}{D^2} \end{aligned}$$

Now we separately consider the estimates of the two expressions:

$$\begin{aligned} & \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right|, \quad \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right|. \\ &= \left| \int_c^d y \left(\bigvee_{j=0}^n (\mu_{A_{k_j}}(x) \cdot \mu_{B_{k_j}}(y)) \right) dy - \sum_{i=0}^n \mu_{A_{k_i}}(x) y_{k_i} \Delta y_{k_i} \right| \\ &= \left| \int_c^d y \left((\mu_{A_{k_s}}(x) \cdot \mu_{B_{k_s}}(y)) \vee (\mu_{A_{k_t}}(x) \cdot \mu_{B_{k_t}}(y)) \right) dy \right. \\ &\quad \left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right|. \end{aligned}$$

1. First we consider $\left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right|$.

By using of the two-phase property of $A_{k_i} (i = 0, 1, \dots, n)$, for $x \in X$, we can easily know the facts as follows:

$$\begin{aligned} & (\exists s, t \in \{0, 1, \dots, n\}) (\mu_{A_{k_s}}(x) + \mu_{A_{k_t}}(x) = 1), \\ & (\forall i \notin \{s, t\}) (\mu_{A_{k_i}}(x) \equiv 0) \end{aligned}$$

Then we have

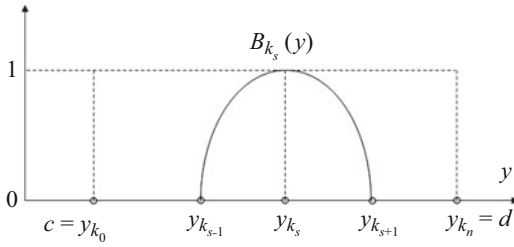
Case 1 $|s - t| = 0$, that is, $s = t$. At this time, there must be $x = x_{k_s}$, that is, x just in the node x_{k_s} , then $\mu_{A_{k_s}}(x) = 1$. We can process the variable y by using Taylor expansion based on three kinds of situations.

Situation 1 $0 < s < n$. Refer to Fig. 2, where

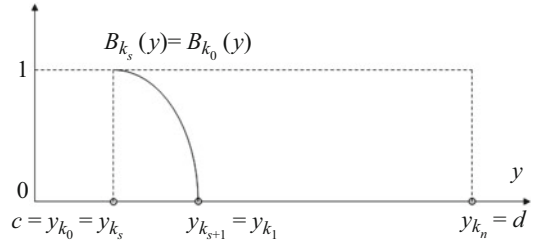
$$B_{k_s}(y) \triangleq \mu_{B_{k_s}}(y).$$

At this situation, there exists $\eta_{s-1} \in (y_{k_{s-1}}, y_{k_s})$ and also exists $\eta_s \in (y_{k_s}, y_{k_{s+1}})$, such that

$$\begin{aligned} & \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| = \left| \int_c^d y \mu_{B_{k_s}}(y) dy - y_{k_s} \Delta y_{k_s} \right| \\ &= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} y \mu_{B_{k_s}}(y) dy + \int_{y_{k_s}}^{y_{k_{s+1}}} y \mu_{B_{k_s}}(y) dy - y_{k_s} \Delta y_{k_s} \right| \\ &= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} \left(y_{k_{s-1}} \mu_{B_{k_s}}(y_{k_{s-1}}) + (y \mu_{B_{k_s}}(y))'_{\eta_{s-1}} (y - y_{k_{s-1}}) \right) dy \right. \\ &\quad \left. + \int_{y_{k_s}}^{y_{k_{s+1}}} \left(y_{k_s} \mu_{B_{k_s}}(y_{k_s}) + (y \mu_{B_{k_s}}(y))'_{\eta_s} (y - y_{k_s}) \right) dy - y_{k_s} \Delta y_{k_s} \right| \\ &= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} (y \mu_{B_{k_s}}(y))'_{\eta_{s-1}} (y - y_{k_{s-1}}) dy + \int_{y_{k_s}}^{y_{k_{s+1}}} (y \mu_{B_{k_s}}(y))'_{\eta_s} (y - y_{k_s}) dy \right| \\ &\leq \int_{y_{k_{s-1}}}^{y_{k_s}} \left| (y \mu_{B_{k_s}}(y))'_{\eta_{s-1}} (y - y_{k_{s-1}}) \right| dy + \int_{y_{k_s}}^{y_{k_{s+1}}} \left| (y \mu_{B_{k_s}}(y))'_{\eta_s} (y - y_{k_s}) \right| dy \\ &\leq \Delta_2^2 \max \left\{ \left| (y \mu_{B_{k_s}}(y))'_{\eta_{s-1}} \right|, \left| (y \mu_{B_{k_s}}(y))'_{\eta_s} \right| \right\} \leq \Delta_2^2 (1 + M_0 K_1) \end{aligned}$$



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 2 $s = t, 0 < s < n$



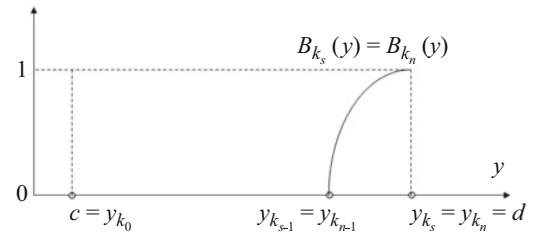
Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 3 $s = t, s = 0$

Situation 2 $s = 0$. Refer to Fig. 3, where

$$B_{k_0}(y) \triangleq \mu_{B_{k_0}}(y).$$

At this situation, there exists $\eta_s \in (y_{k_s}, y_{k_{s+1}})$, such that

$$\left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| = \left| \int_{y_{k_s}}^{y_{k_{s+1}}} y \mu_{B_{k_s}}(y) dy - y_{k_s} \Delta y_{k_s} \right| \leq \frac{\Delta_2^2}{2} (1 + M_0 K_1)$$



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 4 $s = t, s = n$

Situation 3 $s = n$. Refer to Fig. 4, where

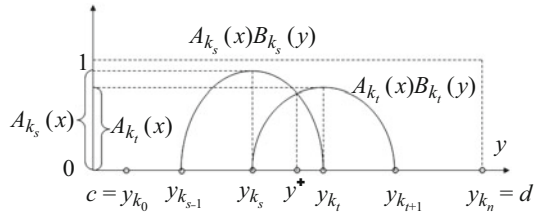
$$B_{k_n}(y) \triangleq \mu_{B_{k_n}}(y),$$

and so on in the following figures. Similarly, we have

$$\left| \int_c^d y R(x, y) dy - P_n(x) \right| \leq \frac{\Delta_2^2}{2} (1 + M_0 K_1)$$

Case 2 $|s - t| = 1$. We only consider that $s < t$, because we can get the same result with $s < t$ or $s > t$. We still process the variable y by using Taylor expansion on three kinds of situations.

(a) $0 < s < t < n$. Refer to Fig. 5. Because all B_{k_i} are fuzzy numbers, there must be a cross point as being $y^* \in (y_{k_s}, y_{k_t})$ between



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 5 $0 < s < t < n$

these membership functions as being $\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y)$ and $\mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y)$ of the fuzzy sets, and then there are four points as the following:

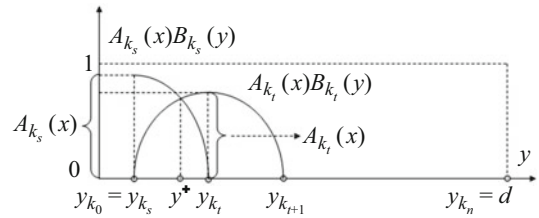
$$\eta_{s-1} \in (y_{k_{s-1}}, y_{k_s}), \quad \eta_1^* \in (y_{k_s}, y^*), \\ \eta_2^* \in (y^*, y_{k_t}), \quad \eta_t \in (y_{k_t}, y_{k_{t+1}})$$

such that

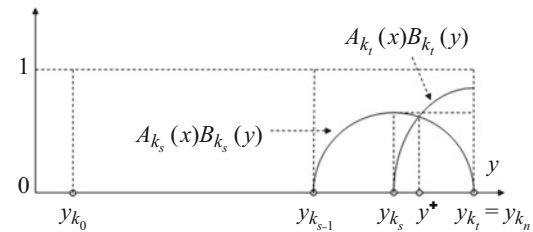
$$\begin{aligned}
& \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| \\
&= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} u \left(\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) \right) dy + \int_{y_{k_s}}^{y^*} \left(\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) \right) dy \right. \\
&\quad \left. + \int_{y^*}^{y_{k_t}} u \left(\mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) \right) dy + \int_{y_{k_t}}^{y_{k_{t+1}}} y \left(\mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) \right) dy \right. \\
&\quad \left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \\
&= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} \mu_{A_{k_s}}(x) \left(y \mu_{B_{k_s}}(y) \right)'_{\eta_{s-1}} (y - y_{k_{s-1}}) dy + \mu_{A_{k_s}}(x) y_{k_s} (y^* - y_{k_s}) \right. \\
&\quad \left. + \int_{y_{k_s}}^{y^*} \mu_{A_{k_s}}(x) \left(y \mu_{B_{k_s}}(y) \right)'_{\eta_1^*} (y - y_{k_{s-1}}) dy + \mu_{A_{k_t}}(x) y_{k_t} (y_{k_t} - y^*) \right. \\
&\quad \left. + \int_{y^*}^{y_{k_t}} \mu_{A_{k_t}}(x) \left(y \mu_{B_{k_t}}(y) \right)'_{\eta_2^*} (y - y^*) dy \right. \\
&\quad \left. + \int_{y_{k_t}}^{y_{k_{t+1}}} \mu_{A_{k_t}}(x) \left(y \mu_{B_{k_t}}(y) \right)'_{\eta_t} (y - y_{k_t}) dy \right. \\
&\quad \left. - \mu_{A_{k_s}}(x) y_{k_s} (y_{k_{s+1}} - y^* + y^* - y_{k_s}) \right| \\
&\leq 2\Delta_2^2(1 + M_0 K_1) + 2M_0 \Delta_2
\end{aligned}$$

(b) $0 = s < t < n$. Similarly, we can have the following inequality (see Fig. 6):

$$\begin{aligned}
& \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| \\
&= \left| \int_{y_{k_s}}^{y^*} y \left(\mu_{A_{k_s}}(x) \cdot \mu_{B_{k_s}}(y) \right) dy \right. \\
&\quad \left. + \int_{y^*}^{y_{k_t}} y \left(\mu_{A_{k_t}}(x) \cdot \mu_{B_{k_t}}(y) \right) dy \right. \\
&\quad \left. + \int_{y_{k_t}}^{y_{k_{t+1}}} y \left(\mu_{A_{k_t}}(x) \cdot \mu_{B_{k_t}}(y) \right) dy \right. \\
&\quad \left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \\
&\leq \frac{3}{2} \Delta_2^2(1 + M_0 K_1) + 2M_0 \Delta_2
\end{aligned}$$

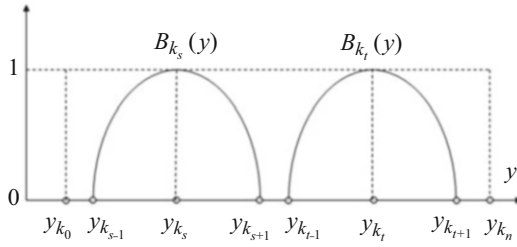


Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 6 $0 = s < t < n$

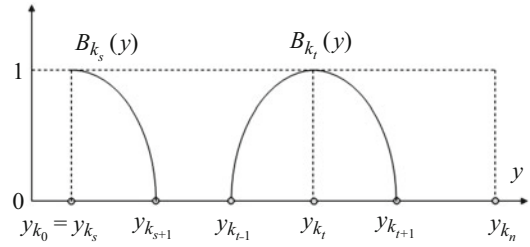


Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 7 $0 < s < t = n$

(c) $0 < s < t = n$. Similar to the former situation, we have the following inequality (see Fig. 7):



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 8 $|s - t| > 1, 0 < s < t < n$



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 9 $|s - t| > 1, 0 = s < t < n$

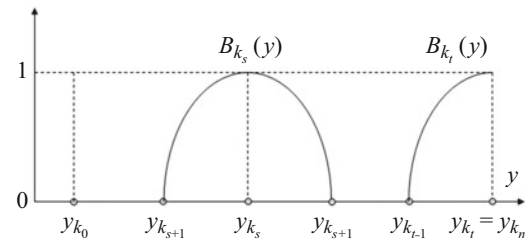
$$\begin{aligned} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| &= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} y (\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y)) dy \right. \\ &+ \int_{y_{k_s}}^{y_{k_t}} y (\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y)) dy + \int_{y_{k_t}}^{y_{k_{t+1}}} y (\mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y)) dy \\ &\left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \\ &\leq \frac{3}{1} \Delta_2^2 (1 + M_0 K_1) + 2M_0 \Delta_2 \end{aligned}$$

Case 3 $|s - t| > 1$. Assume that $s < t$, and the situation for $s < t$ is the same. We can process the variable y by using Taylor expansion on four kinds of situation.

(a) $0 < s < t < n$. Refer to Fig. 8. For this situation, we have the following inequality:

$$\begin{aligned} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| &= \left| \int_{y_{k_{s-1}}}^{y_{k_s}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy \right. \\ &+ \int_{y_{k_s}}^{y_{k_{s+1}}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} \\ &+ \int_{y_{k_{t-1}}}^{y_{k_t}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dy \\ &\left. + \int_{y_{k_t}}^{y_{k_{t+1}}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dy - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \\ &\leq 2\Delta_2^2 (1 + M_0 K_1) \end{aligned}$$

(b) $0 = s < t < n$. It is easy to know the following inequality (see Fig. 9):



Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 10 $|s - t| > 1, 0 < s < t = n$

$$\begin{aligned} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| &= \left| \int_{y_{k_s}}^{y_{k_{s+1}}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy \right. \\ &+ \int_{y_{k_{t-1}}}^{y_{k_t}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dy + \int_{y_{k_t}}^{y_{k_{t+1}}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dt \\ &\left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \leq \frac{3\Delta_2^2}{2} (1 + M_0 K_1) \end{aligned}$$

(c) $0 < s < t = n$. Also we can easily know the following inequality (see Fig. 10):

$$\begin{aligned} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| &= \left| \int_{y_{k_{t-1}}}^{y_{k_t}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy \right. \\ &+ \int_{y_{k_s}}^{y_{k_{s+1}}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy + \int_{y_{k_{t-1}}}^{y_{k_t}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dy \\ &\left. - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \leq \frac{3\Delta_2^2}{2} (1 + M_0 K_1) \end{aligned}$$

(d) $0 = s < t = n$. At last, we can know the following equation:

$$\begin{aligned}
& \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| = \left| \int_{y_{k_s}}^{y_{k_s+1}} y \mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y) dy \right. \\
& + \int_{y_{k_{t-1}}}^{y_{k_t}} y \mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y) dy - \mu_{A_{k_s}}(x) y_{k_s} \Delta y_{k_s} \\
& \left. - \mu_{A_{k_t}}(x) y_{k_t} \Delta y_{k_t} \right| \leq \Delta_2^2 (1 + M_0 K_1)
\end{aligned}$$

For all three kinds of cases, we get the inequality:

$$\begin{aligned}
\left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| & \leq 2\Delta_2^2 (1 + M_0 K_1) \\
& + 2M_0 \Delta_2.
\end{aligned}$$

2) We turn to the estimation of

$$\left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right|.$$

In fact, we also use the two phase property on A_i . For a given point $x \in X$, we know the fact that

$$(\exists s, t \in \{0, 1, \dots, n\}) (\mu_{A_{k_s}}(x) + \mu_{A_{k_t}}(x) = 1).$$

Then we have the following equation:

$$\begin{aligned}
& \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \\
& = \left| \int_c^d \left(\bigvee_{j=0}^n (\mu_{A_{k_j}}(x) \mu_{B_{k_j}}(y)) \right) dy - \sum_{i=0}^n \mu_{A_{k_i}}(x) \Delta y_{k_i} \right| \\
& = \left| \int_c^d \left((\mu_{A_{k_s}}(x) \mu_{B_{k_s}}(y)) \vee (\mu_{A_{k_t}}(x) \mu_{B_{k_t}}(y)) \right) dy \right. \\
& \quad \left. - \mu_{A_{k_s}}(x) \Delta y_{k_s} - \mu_{A_{k_t}}(x) \Delta y_{k_t} \right|
\end{aligned}$$

Case 1 $|s - t| = 0$, that is $s = t$. Similarly, we process the variable y by using Taylor expansion on three kinds of situation, and get the following results:

(a)

$$0 < s < n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq K_1 \Delta_2^2.$$

(b)

$$s = 0 \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq \frac{\Delta_2^2}{2} K_1.$$

(c)

$$s = n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq \frac{\Delta_2^2}{2} K_1.$$

Case 2 $|s - t| = 1$. Assume that $s < t$ and for $s > t$, the situation is the same. We still process the variable y by using Taylor expansion on three kinds of situation.

(a)

$$\begin{aligned}
0 < s < t < n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \\
\leq 2\Delta_2^2 K_1 + 2\Delta_2.
\end{aligned}$$

(b)

$$\begin{aligned}
0 = s < t < n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \\
\leq \frac{3}{2} \Delta_2^2 K_1 + 2\Delta_2.
\end{aligned}$$

(c)

$$\begin{aligned}
0 < s < t = n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \\
\leq \frac{3}{2} \Delta_2^2 K_1 + 2\Delta_2.
\end{aligned}$$

Case 3 $|s - t| > 1$. Assume that $s < t$ and for $s > t$, the situation is the same. We can process the variable y by using Taylor expansion on four kinds of situations as the following:

(a)

$$\begin{aligned}
0 < s < t < n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \\
\leq 2\Delta_2^2 K_1.
\end{aligned}$$

(b)

$$0 = s < t < n \Rightarrow \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| \leq \frac{3\Delta_2^2}{2} K_1.$$

(c)

$$0 < s < t = n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq \frac{3\Delta_2^2}{2} K_1.$$

(d)

$$0 = s < t = n \Rightarrow \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq \Delta_2^2 K_1.$$

For all three kinds of cases, we get the following inequality:

$$\left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right| \leq 2\Delta_2^2 K_1 + 2\Delta_2.$$

In the end, according to the results of (1) and (2), we have

$$\begin{aligned} & |\bar{s}_n(x) - f_n(x)| \\ & \leq \frac{\bar{Q} \left| \int_c^d y \mu_R(x, y) dy - P_n(x) \right| + \bar{P} \left| \int_c^d \mu_R(x, y) dy - Q_n(x) \right|}{D^2} \\ & \leq 2 \frac{2\Delta_2(\Delta_2^2(1 + M_0 K_1) + M_0 \Delta_2) + 2M_0 \Delta_2(\Delta_2^2 K_1 + \Delta_2)}{D^2} \\ & = 4 \frac{2M_0 + 2M_0 K_1 \Delta_2 + \Delta_2}{D^2} \Delta_2^2 \end{aligned}$$

Because of $\Delta_2 \Leftrightarrow c\Delta_1$, we can write

$$\Gamma \triangleq 4 \frac{2M_0 + 2M_0 K_1 N + N}{D^2}.$$

Therefore we have

$$\begin{aligned} \|\bar{r}_n\|_\infty &= \|\bar{s}_n - f_n\|_\infty = \max_{x \in X} \{|\bar{s}_n(x) - f_n(x)|\} \\ &\leq \Gamma \Delta_1^2 \end{aligned}$$

The theorem is completely proved. **Q.E.D.**

Corollary 2 In theorem 3, when all B_i are triangular fuzzy sets, then Γ in (3) turn into the following equation:

$$\Gamma = 4 \frac{2M_0 + 2M_0 N_0 + N}{D^2} \quad (33)$$

where $N_0 = \max_{0 \leq i \leq n-1} \left\{ \frac{\Delta_2}{\Delta_{y_i}} \right\}$. **Q.E.D.**

According to the results of theorem 2 and theorem 3, we can get the conclusion that we need and is shown as the following theorem.

Theorem 4 In the conclusion of theorem 2 and theorem 3, if we write $r_n^*(x) \triangleq s(x) - \bar{s}_n(x)$, then we have the following inequality:

$$\begin{aligned} \|r_n^*\|_\infty &= \|s - \bar{s}_n\|_\infty \leq \|s - f_n\|_\infty + \|f_n - \bar{s}_n\|_\infty \\ &= \|r_n\|_\infty + \|\bar{r}_n\|_\infty \leq (M\Delta_1 + \Gamma)\Delta_1^2 \end{aligned} \quad (34)$$

where the meanings of symbols M, Γ, Δ_1 have been provided in theorem 2 and theorem 3. **Q.E.D.**

Simulation Example

For the function $s(x) = \sin x \in C[-3, 3]$, clearly we know that $\|s'\|_\infty = \|s''\|_\infty = 1$. Consider the approximate effects of $\bar{s}(x)$ to $s(x)$ and $f_n(x)$ to $s(x)$. For simplification, the universe $X = [-3, 3]$ is equidistantly partitioned as

$$\lambda = \Delta x_i = \frac{6}{n}, i = 0, 1, \dots, n-1.$$

Take

$$\mu_{A_{k_i}^*}(x) = \frac{\mu_{A_{k_i}}(x)\Delta y_{k_i}}{\sum_{j=0}^n \mu_{A_{k_j}}(x)\Delta y_{k_j}}, i = 0, 1, \dots, n,$$

$$\bar{s}_n(x) = \frac{\int_0^1 y \mu_{B(\xi=x)}(y) dy}{\int_0^1 \mu_{B(\xi=x)}(y) dy} \approx \frac{\sum_{i=0}^n \mu_{B(\xi=x)}(y_i) y_i \Delta y_i}{\sum_{i=0}^n \mu_{B(\xi=x)}(y_i) \Delta y_i}$$

$$= \sum_{i=0}^n A_i^*(x) y_i = f_n(x),$$

where $A_{k_i}(x)$ ($i = 0, 1, \dots, n$) are taken as triangle fuzzy sets.

The approximate situations to $s(x)$ are shown in Fig. 11, where No.1 Trimf A_i represents triangle

fuzzy set, No.2 Trimf $B_{d(i)}$ represents triangle wave function $B_{d(i)}$, No.3 Contrast1 represents the comparison between curve $s(x)$ and curve $\bar{s}(x)$, No.4 Contrast 2 represents the comparison between curve $s(x)$ and curve $f_n(x)$, and the curve represent sine function, the curve “—” represents $\bar{s}(x)$, the curve “—” represents $f_n(x)$; and the following curves are the same.

When $n = 30$, the approximate situation of $\bar{s}(x)$ and $f_n(x)$ to $s(x)$ are shown in Fig. 12.

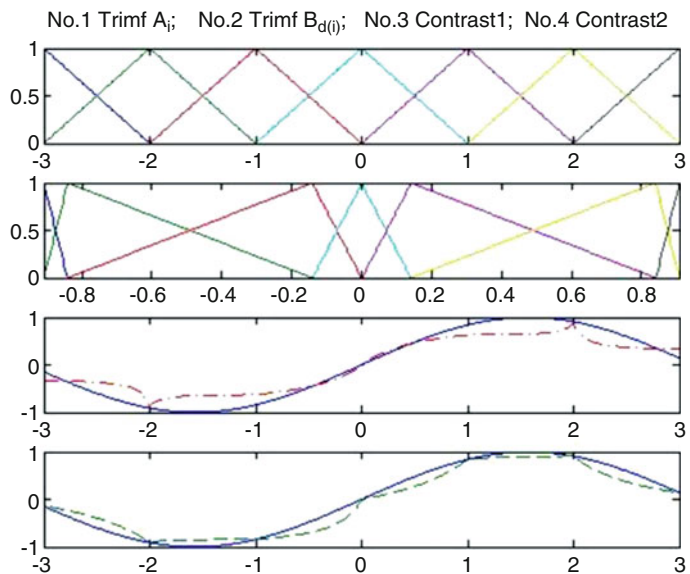
When $n = 6$, the approximate accuracy of $\bar{s}(x)$ and $f_n(x)$ to $s(x)$, respectively, are 0.3561 and 0.2780; While $n = 30$, the approximate accuracy of $\bar{s}(x)$ and $f_n(x)$ to $s(x)$, respectively, are 0.0329 and 0.0443.

From the figures, we can see that when $n = 6$ (that is, there are 7 rules) the approximate accuracy is low, but the curve has a better smoothness; while there are 31 rules, the approximate accuracy is higher, and the smoothness is better. So at this situation, it's better to approximate sine function $s(x)$ by using curve $\bar{s}(x)$.

Conclusions

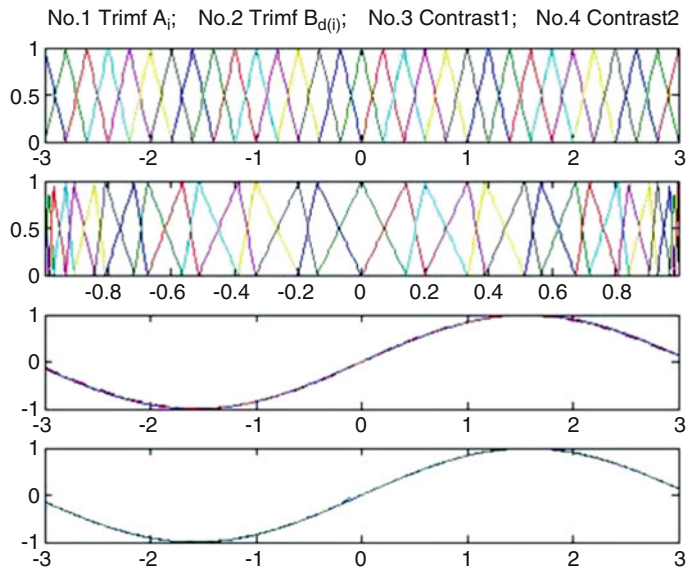
In this entry, function approximation effect analysis of fuzzy systems is researched in detail. First, the structures of fuzzy systems are discussed

Function Approximation Property of Fuzzy Systems and Its Error Analysis, Fig. 11 $n = 6$, approximation of $\bar{s}(x)$ and $f_n(x)$ to $s(x)$



Function Approximation Property of Fuzzy Systems and Its Error

Analysis, Fig. 12 $n = 30$,
approximation of $\bar{s}(x)$ and
 $f_n(x)$ to $s(x)$



where a kind of important structure is formed by the so-called CRI method and the gravity method defuzzification and Larsen implication operator. Second, it is proved that this kind of fuzzy systems can approximate any real continuous function to arbitrary accuracy in a general sense. Third, the error estimate of the function approximation is given. Finally, simulation example results show that the fuzzy systems proposed in this entry has not only good approximation accuracy but also a smooth approximation curve.

Function approximation property and approximation effect analysis of fuzzy systems are essentially belonging to function approximation theory in real analysis or mathematical analysis. So we should adequately use classical tools of real analysis or mathematical analysis to study fuzzy systems.

Bibliography

- Buckley JJ (1993) Sugeno type controllers are universal controllers. *Fuzzy Sets Syst* 53:299–304
 Castro JL (1995) Fuzzy logic controllers are universal approximators. *IEEE Trans SMC* 25:629–635

- Li HX (1998) Interpolation mechanism of fuzzy control. *Sci China (Series E)* 41(3):312–320
 Li HX (2006) Probability representations of fuzzy systems. *China Sci (F Series)* 49(3):339–363
 Li YM, Shi Z-K, Li Z-H (2002a) Approximation theory of fuzzy systems based upon genuine many-valued implications—SISO cases. *Fuzzy Sets Syst* 130:147–157
 Li YM, Shi ZK, Li ZH (2002b) Approximation theory of fuzzy systems based upon genuine many-valued implications—MIMO cases. *Fuzzy Sets Syst* 130:159–174
 Liu P, Li H (2001) Analyses for π -norm approximation capability of Mamdani fuzzy systems. *Inf Sci* 138:195–210
 Liu P, Li H (2005) Approximation of stochastic processes by T–S fuzzy systems. *Fuzzy Sets Syst* 155:215–235
 Tikk D, Koczy LT, Gedeon TD (2003) A survey on universal approximation and its limits in soft computing techniques. *Int J Approx Reason* 33:185–202
 Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans SMC* 3:28–44
 Zeng X-J, Keane JA (2005) Approximation capabilities of hierarchical fuzzy systems. *IEEE Trans Fuzzy Syst* 13(5):659–672
 Zeng X-J, Singh MG (2000) Approximation accuracy analysis of fuzzy systems as function approximators. *IEEE Trans Fuzzy Syst* 4(1):44–59
 Zeng K, Zhang N-Y, Xu W-L (2000) A comparative study on sufficient conditions for Takagi–Sugeno fuzzy systems as universal approximators. *IEEE Trans Fuzzy Syst* 8(6):773–780



Issues in Access Control and Privacy for Big Data

Sylvia L. Osborn, Daniel Servos and
Motahera Shermin
Department of Computer Science, Western
University, London, ON, Canada

Article Outline

Glossary
Definition of the Subject
Introduction
Access Control and Privacy
Access Control
Privacy
Access Control Versus Privacy
4 Vs of Big Data
Volume
Variety
Velocity
Veracity
Putting it All Together
Summary and Future Directions
Bibliography

Glossary

Access Control The selective restriction of access to a physical or virtual resource under some predefined model or set of policies.

Attribute-Based Access Control (ABAC) A policy-based access control model in which access control decisions are made based on the attributes of the users, resources, and the system's environment.

Big Data The large-scale collection, processing, and analysis of diverse sets of information from rapidly growing sources.

Discretionary Access Control (DAC) An access control model that allows individual

users to govern access to objects and resources they own by granting or revoking specific permissions to other users or groups.

Mandatory Access Control (MAC) A label-based model of access control where access is decided based on comparing the label of the subject (clearance) with the label (security level) of the object being accessed.

Privacy The protection and proper handling of data with respect to sharing with third parties, conforming to regulatory obligations, and handling user consent to information collection and storage.

Role-Based Access Control (RBAC) A model of access control in which permissions are granted based on the organizational roles assigned to users.

Definition of the Subject

This entry explores issues in access control and privacy for big data. We highlight the differences we see between access control and privacy. We examine the requirements introduced by having big data, specifically by the Vs of big data, rather than data that can be dealt with on a simple platform. Then we look at the additional challenges introduced when this big data is housed in a cloud.

Introduction

With the advent of massive storage and massively parallel computation, coupled with the increasing adoption of internet platforms and applications, we are now in the era of big data. *Big Data* is generally considered to encompass massive collections of data and the processing/analysis required on it. The data is so massive that traditional relational databases or desktop analytical tools are not adequate to perform the capture, or required analysis, with acceptable speed. The processing envisaged varies widely from traditional integration and querying tasks to complex

S. L. Osborn: deceased.

© Springer Science+Business Media, LLC, part of Springer Nature 2023
T.-Y. Lin et al. (eds.), *Granular, Fuzzy, and Soft Computing*,
https://doi.org/10.1007/978-1-0716-2628-3_752

Originally published in
R. A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*, © Springer Science+Business Media LLC 2021
https://doi.org/10.1007/978-3-642-27737-5_752-1

analyses and visualizations. Sources of big data are varied, ranging from widely deployed sensors to commercial websites and social media.

Since this big data is probably not under the control of a traditional database package, there is not an obvious way to specify access control restrictions on it, and no traditional way to ensure privacy requirements are met. Both access control and privacy protection are required to protect data from unwanted access and breaches of privacy. Having these concerns dealt with in individual applications is not adequate. We are reminded of the days before database systems when multiple copies of the same data might exist within a company, and each application had its own version of the schema and its own copy of the data. Database packages brought the schema out of the individual applications so that a single version of the description of the data (the schema), and the data itself, could be managed independently of various applications. Having protection for a big data repository, independent of the applications accessing it, is the goal of access control and privacy protection for big data. This goal has not yet been achieved.

We begin with a review of what we mean by access control and privacy, and how they differ. We then discuss the four Vs of big data and what requirements they raise as far as achieving access control and privacy protection. Lastly we bring in the cloud, provide a summary and discuss future directions.

Access Control and Privacy

Access Control

Access control in traditional operating system and database environments deals with which users can perform which operations on what data. The traditional access control models are Discretionary Access Control (DAC), Mandatory Access Control (MAC), and Role-based Access Control (RBAC). Discretionary access control is based on the concept of ownership of data: each user can, at their discretion, give access permissions to other users to access data which they own. This model has inherently decentralized control in that every user is making access control decisions

concerning their own data, and it is difficult to formulate a universal picture of what is going on. DAC is provided by relational database packages and its administrative GRANT and REVOKE commands are part of the SQL standard (ISO 2011). Although DAC is inherently decentralized, it is possible with a relational database system to assume that the database administrator owns all the data, and thus the database administrator can use the DAC commands to design access control for applications such that the design is, in fact, centrally administrated.

In MAC, all objects and subjects (users/applications) have a label, and access (usually restricted to read and write) is decided by comparing the label of the subject (sometimes referred to as its clearance) with the label (security level) of the object they are trying to access (Landwehr 1981; Sandhu 1993). MAC is generally considered to be too rigid for commercial applications.

In RBAC (Sandhu et al. 1996), permissions are gathered into roles, and roles are assigned to users. It is centrally managed and designed, and can be configured to simulate DAC and MAC (Osborn et al. 2000). Designing a system using roles is called role engineering, and can be achieved either both bottom-up or top-down (a good introduction can be found in the work by Coyne and Davis (2007)). Bottom-up design works from existing user-permission assignments to construct roles, whereas top-down design proceeds from a security designer's understanding of the job functions in a company or the requirements of various types of users in a software application. Top-down design is performed by humans and is considered slow and expensive. With bottom-up design, sometimes called role mining, it is difficult to attach semantics (or even a role name) to the results (Molloy et al. 2010). Because roles are at a coarser granularity than individual users and individual permissions, RBAC models are considered to be more efficient to manage, and easier to design. However, since an RBAC system has to be centrally designed, it may not be agile enough for a quickly changing, unpredictable application area.

More recently, Attribute-based Access Control (ABAC) has been proposed (Servos and Osborn

2017; Jin et al. 2012; Servos and Osborn 2014). With attribute-based access control, users acquire attributes through various means, which are then matched with object attributes to decide whether access should be allowed. It is inherently fine-grained, and allows a decentralized administration of security policy.

Once an access control system has been designed and deployed, its implementation usually includes a reference monitor, the part of a system through which all access requests must pass. One technique for implementing a reference monitor is to recode the DAC or RBAC policy as rules in XACML, and have an XACML engine act as the reference monitor (Anderson 2004; Ferrini and Bertino 2009).

Privacy

Access control is used to manage all kinds of data; data about inventory, weather patterns, ocean temperatures as well as bank accounts and online shopping habits. Considerations of privacy deal with “sensitive” information relating to people, or perhaps organizations. Many countries and jurisdictions have regulations governing the privacy of data concerning individuals stored in online systems. Whereas access control for databases usually protects facts as they exist in the database, privacy concerns can require the protection of facts, summaries, or partial information (Adam and Worthmann 1989; Barker et al. 2009). Barker et al. (2009) also point out that when, for example, a customer of an on-line shopping site gives some of their information to the shopping site, they no longer “own” it in the customary sense of DAC. It is now owned by the shopping site, which they refer to as the “house.”

Dwork (2008) has used the term *curator* to refer to the process that releases information from a statistical database. Statistical databases do not allow retrieval of facts, but only allow retrieval of statistical information like averages, frequency counts, etc. In general, the curator may: (a) limit what queries are answered, (b) may limit what data is used to compute the answers or (c) may introduce “falsehoods” to comply with the privacy requirements of the statistical database. For case (a), consider a relational database

tracking medical diagnoses and treatments within a hospital. Some columns, such as the name or date of birth of a patient, would be used for regular administration of the hospital, whereas other columns, the diagnosis of a patient would be a prime example, might be the subject of statistical operations for research purposes. This column would also be considered one that must be kept private. Only statistical summaries such as a frequency count would be allowed to be released for this column. Queries can be posed which read other columns but only report statistics (such as counting the frequency of each diagnosis for female patients over 65). Limiting queries can be done by limiting the size of the tuple subset that is used to compute any answer to a query, so that it never, for example, is the result of summarizing k tuples or $n - k$ tuples, where n is the size of the table and k is quite small. Any query which involves too few, or too near the total number of tuples, is simply not answered. Even with this, Denning has shown that, using something called a tracker, and asking several well-constructed queries, individual data can still be revealed (Denning and Denning 1979).

An example of (b), limiting what data is used to compute the answer to a query, would be using random samples before calculating the statistical quantity. An example of (c), introducing “falsehoods,” is computing random perturbations of the raw data before calculating the statistics. These last two techniques may result in poor quality of the resulting statistics. A summary of the trade-offs in querying statistical databases is given by Adam and Worthmann (1989).

Further to the above discussions of statistical databases is the concept of *k-anonymity* (Sweeney 2002), which says that statistics about data should not be released unless the information for each person contained in the release cannot be distinguished from at least $k - 1$ other individuals. Additionally, any group of records which has *k-anonymity* can also be required to exhibit *l-diversity*, which means that the group contains l “well represented” values for sensitive attributes (Machanavajjhala et al. 2007). A more recently proposed concept, called *differential privacy* (Dwork 2008) says, informally, that the presence

or absence of a single element will not affect the output of a query in a significant way. This again refers to release of (statistical) information from a statistical database.

Not only are statistical applications able to infer sensitive information about people that they would prefer to keep secret, many current systems track users’ behaviors, say by tracking their movements through their cell phone, or tracking their interests by monitoring the websites they visit. This data obtained through monitoring is data the user has no idea exists, so they are not able to express preferences about this type of data. Another example is a large database of medical information, which might be used for research. The data is definitely no longer under any control of the patients who are described in it. Controls are needed so that inferences made from the data cannot be used to release private information.

Still relating to privacy, models have been introduced (such as the model by Byun and Li (2008)) to allow individual users to attach an intended purpose to parts of data about them stored in a database, which is no longer under their control (the “house” situation from Barker et al. (2009)). Using these purpose specifications, if a customer’s data is stored on the computers of an online shopping website, for example, the customer could indicate that their address can be used for the shipping purpose but not for marketing. In Byun and Li (2008), examples show that there are cases where the labels should be attached to individual attribute values, to tuples, to columns, or to whole tables. A discussion of trade-offs in implementing attribute labels is given by Mahmud and Osborn (2012).

Access Control Versus Privacy

Access control is a straightforward protection of facts in a database and controlling who can access and manipulate these facts. When the facts pertain to data that is considered sensitive for some reason, this protection has been called privacy protection. Some contrasts between what we mean by access control and privacy are highlighted in Table 1. The first point emphasizes that privacy

Issues in Access Control and Privacy for Big Data,
Table 1 Differences between Access Control and Privacy Protection

	Access control	Privacy
1	Who can access what in what manner	“Sensitive” information with extra protection requirements
2	Data can be about anything	Data is about a person or an organization
3	Usually protecting facts	Can concern specific facts, partial information, or existence of information; inferences
4	No purpose specification	Purpose can be attached at arbitrary granularity
5	In DAC, decisions are individual, MAC and RBAC are centrally managed	Decisions can be individual

concerns add the idea that some information is sensitive to what would otherwise be considered access control. The second point suggests that data that is considered “sensitive” usually concerns people or human organizations, whereas things like massive sensor data, say ocean temperatures or weather data, would not be considered sensitive and that for this data, traditional access control mechanisms should be adequate. The third point suggests that for systems that can infer information from the raw data, such as statistical databases or some online tracking of one’s web browsing habits, then extra provisions beyond traditional access control are needed. Point 4 suggests that if a system is to allow data providers to give additional privacy preferences in the form of purposes for the use of the data about them, then extra mechanisms beyond traditional access control are needed. The final point deals with how centralized/decentralized the administration of access control and privacy protection (ACPP) is. In an online social network with millions of users, allowing each user to specify their access control facts is a huge challenge. If privacy purposes are allowed to be given, then a

mechanism to adhere to these requirements of users can be developed, but checking these preferences in a very big database, be they just access control permissions or privacy purpose labels, can still be very challenging simply because of the volume of data.

One of the most commonly used types of applications today is online social networks. In an online social network, protection from disclosure about some of the data concerning an individual is usually at the fact level, no inferences or statistical summaries are being calculated. Within these systems, deciding how to protect ones' data is popularly referred to as a privacy setting. Since it is at the fact level, access control technologies can be used directly. However, the decisions vary widely from individual to individual, so that they present a highly decentralized challenge for traditional access control methods.

Some newer access control models blur the lines between access control and privacy. P-RBAC extends the RBAC model to support complex privacy policies that include purposes and obligations (Ni et al. 2010). Kolter et al. (2007) show how to use ABAC to incorporate the privacy preferences of the users in access control decisions.

4 Vs of Big Data

In this section, we will discuss 4 Vs of big data. The basic Vs, Volume, Variety and Velocity, are always part of a discussion of big data (Laney 2001; Hurt 2012), so these can be considered as basic Vs. We also talk about Veracity, as it seems to be becoming an essential *V*. Others have proposed Value as a *V* (see, e.g., the work by Demchenko et al. (2013)), but we do not include value here, as all data stored can be assumed to have some value, and this does not change because it is big. A diagram of the three basic Vs, showing how the characteristics of the data change as we move from small, centralized systems to cloud systems, can be found in the article by Hurt (2012). We have adapted this diagram to include Veracity (see Fig. 1), and show some

of the characteristics these 4Vs introduce, as we move out from the origin, which represents a centralized system. It should be noted that properties mentioned in one ring in one quadrant do not necessarily go together with those in the same ring in another quadrant.

As we describe each *V*, we will also discuss what we consider to be the challenges to access control and privacy protection (ACPP) posed by this *V*.

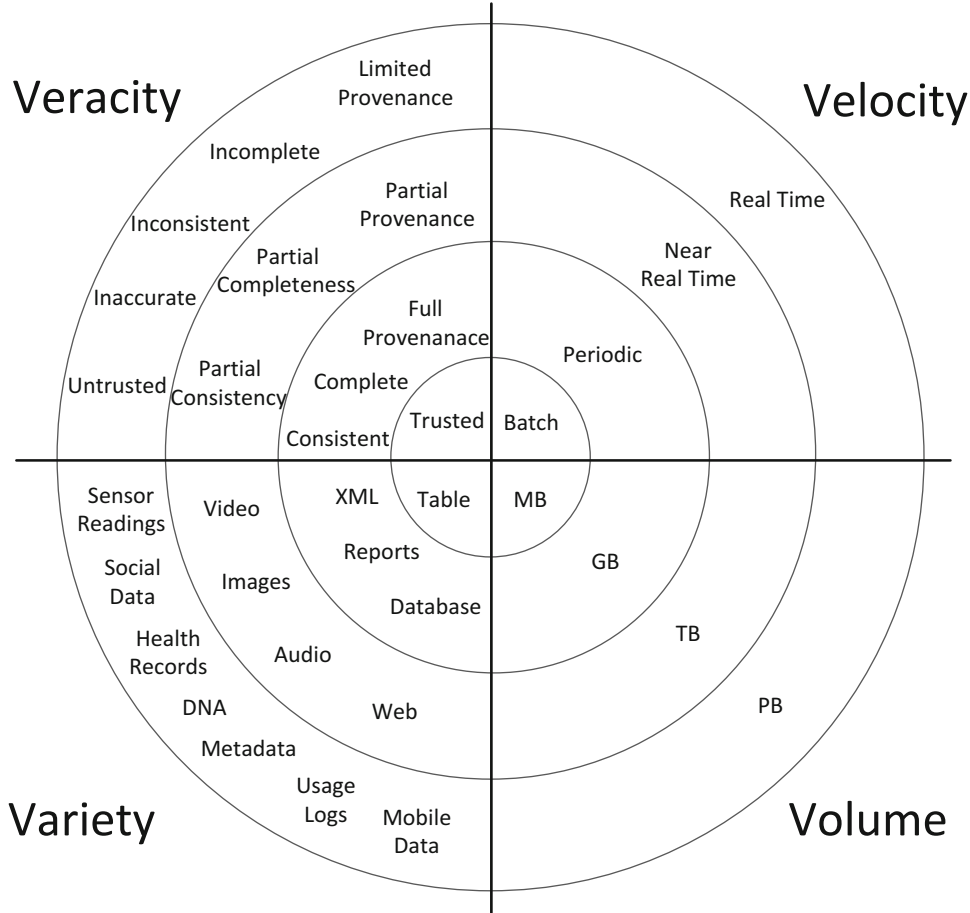
Volume

With the increasing connectedness of the world, the amount of data available for analytics is constantly increasing.

Whereas massive volumes of data pose problems for storage, processing, and analytics of the data, we do not believe they add any specific requirements for ACPP, other than to say that the implementation of the ACPP mechanisms needs to be very efficient. The huge volume of data implies that the platform for the applications will not be a traditional system but probably some form of distributed architecture, commonly leveraging the cloud service model. Issues introduced by this platform will be discussed in the next section.

Variety

Variety indicates that the kinds of data vary widely, from traditional tuple-like database data to text, images, documents, streams of readings from sensors, results from DNA analyses, key-value pairs, etc. Within a single, large application, say dealing with many streams of sensor readings, there might not be much variety in the types of data, but others, e.g. a news stream with images, blogs, text summaries, video etc., would have large variety to deal with. It is also unlikely that the same software package will deal with both of these examples, so rather than having a business world where everyone is using relational databases, we now have a world in which there is a large variety of software dealing with massive amounts of data. Variety also suggests a large variety of the so-called privacy requirements when a large, online social network has millions of users.



Issues in Access Control and Privacy for Big Data, Fig. 1 Characteristics Introduced by Increases in the 4Vs of Big Data

Because of the wide variety of data and processing, ACPP mechanisms need to be available beyond what is provided by traditional database systems. Much of the data will be managed by other systems that are not traditional databases, so these systems need to have ACPP mechanisms themselves, or a way of providing ACPP outside of these applications needs to be available. To deal with a Facebook-like environment, a much decentralized access control model is required. These models need to be agile, flexible, and capable of dealing with a large variety of requirements.

Velocity

Velocity as a characteristic of big data means that in some environments, the data arrives very

quickly, possibly by some streaming interface. Velocity also suggests that data types and usage patterns can change very quickly.

Considerations of access control for streaming data are not well developed. Some work on privacy for streaming data has been done (e.g., Cao et al. 2010; Zakerzadeh and Osborn 2013). A centralized design of the access control model may not work well if data and usage patterns change quickly; quickly changing requirements mean the AACPP mechanisms need to be agile. Velocity also requires that the implementation of the ACPP mechanisms needs to be very efficient.

Veracity

Veracity indicates a high level of data quality, also sometimes referred to as provenance, is required. With massive amounts of data arriving at a data center, knowing that the data is relatively free of error can be an important requirement.

Traditionally in a database system, ensuring that the data is updated by qualified individuals can be ensured by restricting who can write to and update the database. This is part of an access control system. The Biba model also speaks to this (Sandhu 1993). Others focus on the sequence of operations that have been performed on data, and its possible migration from one platform to another. Through all of these steps, trusted agents need to be handling the data to ensure its integrity. This history of the data is called its provenance. To ensure data quality, provenance and auditing of these activities is important.

Putting it All Together

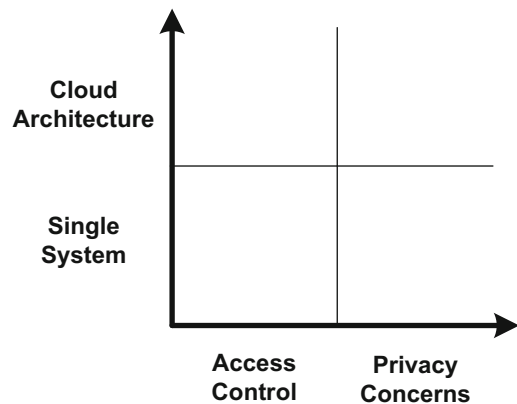
Big data implies that the data is too big to fit on a traditional computing platform, and that it probably resides in a cloud. NIST defines a cloud as a “model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (Mell et al. 2011). Clouds can be private, community, public, or hybrid clouds. Private clouds serve a single organization and are no different from traditional computing systems in their consideration of ACPP. Community, public, and hybrid clouds may store data in a location not under control of the consumer, and be managed and controlled by entities outside of the organization that provided the data.

As documented by Mell et al. (2011), cloud computing environments are available with different service models. The most bare-bones is Infrastructure as a Service (IaaS), which just provides machines, storage, and networks and requires the customer to load all their software, including the operating systems and applications. Platform as a

Service (PaaS) cloud providers give the customer operating systems on top of the bare-bones cloud, but the customer must provide their own applications. Software as a Service (SaaS) clouds provide other software configurations such as database packages. Many new software packages aimed at handling large amounts of data, which are not necessarily relational, the so-called NoSQL databases, are focused on speed and availability, with less focus on transactional consistency and access control (Grolinger et al. 2013).

Consider the grid shown in Fig. 2. The bottom, left quadrant represents traditional access control on a single system. This is well understood, and handled by traditional database management systems. What happens when we move from the bottom, left quadrant of the figure to the other quadrants? Moving to the bottom, right quadrant means that we are more focused on privacy, still with a single system. We have more concern about the types of inference that may take place, and the labeling of parts of the data for different purposes. In some contexts we move from having a reference monitor to having a curator or auditor. The previously discussed mechanisms for handling data privacy go a long way toward satisfying concerns for this quadrant.

As we move from a single system to a private cloud, we observe that this is basically a distributed database system under the control of a single entity, for which ACPP should be achievable



Issues in Access Control and Privacy for Big Data, Fig. 2 Moving From Traditional Database Access Control to Big Data Access Control and Privacy

using well-known techniques. In moving to a non-private cloud, many new concerns are added. Many issues are introduced because the data is now outsourced. Rather than using a commercial relational database package that offers some straightforward access control, the application may now be running on one of the NoSQL platforms, which have very little ACPP protection (Shermin 2013). We no longer know who all the users are who may be threatening our data. Cloud systems often replicate data for performance reasons. Data may cross international boundaries and be subjected to different regulations. Some mechanisms are available for this quadrant. For example, Amazon Simple Storage Service provides access control lists for data in a simple storage system (Amazon 2020). Office 365 and App fabric are, respectively, the SaaS and PaaS platforms from Microsoft's cloud solution called Azure (Microsoft 2020). Azure's access control system provides many mechanisms such as identity management, a rule engine, and RBAC.

Moving on to the top right quadrant, the challenges are much greater. Some privacy issues are handled by access control mechanisms; if these exist for access control in cloud environments they can be applied directly to deal with these privacy concerns. Solutions dealing with things like *k-anonymity* and differential privacy can be deployed at the point of the curator in a cloud environment. Most recent publications (Colombo and Ferrari (2015) offer a good example) examined in this quadrant give lists of issues but few solutions. Cavoukian et al. (2015), on the other hand, strongly urge practitioners to use ABAC to ensure privacy for big data.

Summary and Future Directions

In this entry we have discussed what we mean by access control and privacy protection, and compared these two requirements to highlight their differences. We have considered the four Vs of big data, and how they add requirements to be satisfied by ACPP solutions.

Summarizing the requirements highlighted by the Vs, solutions for big data need to be agile, efficient, flexible, and need to scale well to very large systems. Privacy concerns suggest that the design/control of ACPP will be decentralized. Variety suggests that many different software platforms might be involved to handle a wide variety of situations, so we can no longer rely on a single technology, such as a relational database package, to handle access control for us. Due to Velocity and Variety, models might be very complicated. Velocity suggests that developers might need to rush to a solution without incorporating mechanisms for access control or privacy.

We then looked briefly at some issues arising when data moves to a cloud computing platform. Some solutions for access control in the cloud have been developed, but mechanisms to ensure privacy in all its dimensions for very big data repositories are still lacking. New models such as ABAC are flexible and more agile than traditional access control models, but it is not yet clear that they can be made efficient for a complicated situation.

Bibliography

- Adam NR, Worthmann JC (1989) Security-control methods for statistical databases: a comparative study. *ACM Comput Surv (CSUR)* 21(4):515–556
- Amazon (2020) Amazon Simple Storage Service (S3). <https://aws.amazon.com/s3/>. Accessed 23 Jan 2020
- Anderson A (2004) XACML profile for role based access control (RBAC). OASIS Access Control TC Committee Draft 1:13
- Barker K, Askari M, Banerjee M, Ghazinour K, Mackas B, Majedi M, Pun S, Williams A (2009) A data privacy taxonomy. In: *British National Conference on Databases*. Springer, pp 42–54
- Byun J-W, Li N (2008) Purpose based access control for privacy protection in relational database systems. *VLDB J* 17(4):603–619
- Cao J, Carminati B, Ferrari E, Tan K-L (2010) Castle: continuously anonymizing data streams. *IEEE Trans Depend Secure Comput* 8(3):337–352
- Cavoukian A, Chibba M, Williamson G, Ferguson A (2015) The importance of ABAC: attribute-based access control to big data: privacy and context. *Privacy and Big Data Institute*, Ryerson University, Toronto
- Colombo P, Ferrari E (2015) Privacy aware access control for big data: a research roadmap. *Big Data Res* 2(4): 145–154

- Coyne EJ, Davis JM (2007) Role engineering for enterprise security management. Artech House, Inc. 685 Canton St. Norwood, MA, United States
- Demchenko Y, Ngo C, de Laat C, Membrey P, Gordijenko D (2013) Big security for big data: addressing security challenges for the big data infrastructure. In: Workshop on Secure Data Management. Springer, pp 76–94
- Denning DE, Denning PJ (1979) The tracker: a threat to statistical database security. *ACM Trans Database Syst (TODS)* 4(1):76–96
- Dwork C (2008) Differential privacy: a survey of results. In: International conference on theory and applications of models of computation. Springer, pp 1–19
- Ferrini R, Bertino E (2009) Supporting RBAC with XACML + OWL. In: Proceedings of the 14th ACM symposium on Access control models and technologies, pp 145–154
- Grolinger K, Higashino WA, Tiwari A, Capretz MA (2013) Data management in cloud environments: NoSQL and NewSQL data stores. *J Cloud Comput Adv Syst Appl* 2(1):22
- Hurt J (2012) The three vs of big data as applied to conferences. <https://velvetchainsaw.com/2012/07/20/three-vs-of-big-data-asapplied-conferences/>
- ISO (2011) ISO/IEC 9075-1-4:2011, information technology-database languages-SQL, part 1–4
- Jin X, Krishnan R, Sandhu R (2012) A unified attribute-based access control model covering DAC, MAC and RBAC. In: IFIP Annual conference on data and applications security and privacy. Springer, pp 41–55
- Kolter J, Schillinger R, Pernul G (2007) A privacy-enhanced attribute-based access control system. In: IFIP Annual conference on data and applications security and privacy. Springer, pp 129–143
- Landwehr CE (1981) Formal models for computer security. *ACM Comput Surv (CSUR)* 13(3):247–278
- Laney D (2001) 3D data management: controlling data volume, velocity and variety. *META Group Res Note* 6(70):1
- Machanavajjhala A, Kifer D, Gehrke J, Venkatasubramanian M (2007) l-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Discov Data (TKDD)* 1(1):3–es
- Mahmud MS, Osborn SL (2012) Tradeoff analysis of relational database storage of privacy preferences. In: Workshop on Secure Data Management. Springer, pp 1–13
- Mell P, Grance T et al (2011) The NIST definition of cloud computing (draft), NIST Spec. Publ. 800:7
- Microsoft (2020) Microsoft Azure Documentation. <https://docs.microsoft.com/en-us/azure/>. Accessed 23 Jan 2020
- Molloy I, Chen H, Li T, Wang Q, Li N, Bertino E, Calo S, Lobo J (2010) Mining roles with multiple objectives. *ACM Trans Inf Syst Secur (TISSEC)* 13(4):1–35
- Ni Q, Bertino E, Lobo J, Brodie C, Karat C-M, Karat J, Trombeta A (2010) Privacy-aware role-based access control. *ACM Trans Inf Syst Secur (TISSEC)* 13(3):1–31
- Osborn S, Sandhu R, Munawar Q (2000) Configuring role-based access control to enforce mandatory and discretionary access control policies. *ACM Trans Inf Syst Secur (TISSEC)* 3(2):85–106
- Sandhu RS (1993) Lattice-based access control models. *Computer* 26(11):9–19
- Sandhu RS, Coyne EJ, Feinstein HL, Youman CE (1996) Role-based access control models. *Computer* 29(2):38–47
- Servos D, Osborn SL (2014) HGABAC: towards a formal model of hierarchical attribute-based access control. In: International symposium on foundations and practice of security. Springer, pp 187–204
- Servos D, Osborn SL (2017) Current research and open problems in attribute-based access control. *ACM Comput Surv (CSUR)* 49(4):1–45
- Shermin M (2013) An access control model for NoSQL databases. Master's thesis, The University of Western Ontario. Electronic Thesis and Dissertation Repository. Paper 1797
- Sweeney L (2002) k-anonymity: a model for protecting privacy. *Int J Uncertainty Fuzziness Knowl Based Syst* 10(05):557–570
- Zakerzadeh H, Osborn SL (2013) Delay-sensitive approaches for anonymizing numerical streaming data. *Int J Inf Secur* 12(5):423–437

Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment

Thomas H. Hinke¹ and Derek Shaw²

¹Moffett Field, CA, USA

²NASA Advanced Supercomputing Division,
NASA Ames Research Center, Moffett Field,
CA, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Security-Relevant Data That Is Useful for

Protecting a Data Center

Risk-Partitioning the Mountain of Data

Data Enrichment to Facilitate Actionable Security

Event Detection

Detecting Actionable Security Events

Future Directions

Bibliography

Glossary

Access control list (ACL) Lists the users or IP addresses that are allowed to access a system.

Advanced Persistent Threat (APT) Malware that has been surreptitiously implanted into a computer system and hides itself over long periods of time in order to exfiltrate valuable data.

Common Vulnerability and Exposure (CVE)

A unique number that identifies a known vulnerability for a computer system.

Dynamic Host Configuration Protocol

(DHCP) A network management protocol

that will dynamically grant IP addresses to systems for their use.

Domain Name System (DNS) Maps a name of a network site into an IP address.

Beacon Communication between malware and command and control (C2) server that typically occurs at a predetermined asynchronous time interval.

Domain An administrative network grouping of subordinate networks or hosts, such as the .edu domain for US universities.

False positive When a system indicates that a particular condition exists, when it does not.

False negative When a system indicates that a particular condition does not exist, when it does.

Intrusion detection system (IDS) A system that monitors computers and networks to detect signs of suspicious activity and generates alerts when such activity is detected.

IP address A set of numbers that provides a unique address to a network system.

Packet inspection A method of examining the contents of network packets that transverse a network. Typically used to identify, classify, and reroute packets based on specific data or payload content.

Stateful firewall Tracks the state of the network connection.

Sudo A Unix/Linux operating system command that allows the user to run programs with the privilege of another user, including the root privilege.

Two-factor authentication An authentication method that requires two different factors to prove identity. Typically, something one has, such as a hardware token, and something one knows, such as a passcode.

Vulnerability scanner Scans a system's software to detect whether it has any of a large set of known vulnerabilities.

Zero-day A vulnerability that is unknown by those who should be interested in mitigating the vulnerability.

Definition of the Subject

This entry describes the security approaches appropriate to a computer data center that is comprised of multiple computers running a range of different operating systems and on which multiple remote users perform a wide variety of processing tasks.

Introduction

In a world where cyber-attacks are ubiquitous, it is important for an organization that maintains data centers not only to harden its defense against such attacks but also to determine which attacks must be mitigated with manual or automated actions. Historically, defenses against such attacks have evolved to keep abreast with the sophistication of the attacks. There are commercial solutions, such as Security Information and Event Management (SIEM) products, that have been developed and deployed to perform some of the information gathering, management, and security actions necessary to protect major data centers. However, this entry instead presents effective alternative approaches to data center defense that are independent of any particular SIEM product.

An organization requires *actionable information* in order to determine what actions, with what timing, are needed to mitigate an attack. Unfortunately, in today's cyber world, such actionable information may be buried in the vast mountain of data that flows into and out of internet-accessible processing centers, including those that support big data and high-performance computing. Ironically, big data techniques may be required to cull actionable data from this mountain of information – the proverbial needle in the haystack.

This entry focuses on approaches that are used in security situational awareness systems (SSAS) to identify actionable security information among an organization's data flows. The canonical organization of interest is a data center that contains a

heterogeneous set of high performance computing systems which process large data sets for a variety of customers. Specifically, this entry examines the following:

- The security-relevant data used to defend the data center against attack
- The partitioning of security-relevant data based on potential indications of actionable security events
- The enrichment of partitioned data through preprocessing to facilitate the detection of actionable security events
- The types of actionable security events that should be detected within the enriched data
- Future directions in security-relevant data processing and analysis

Security-Relevant Data That Is Useful for Protecting a Data Center

Security-relevant data is a broad class of data from a variety of sources that contains information which can facilitate the defense of a data center against attack.

Vulnerability Data

Vulnerability data is information pertaining to a system's weakness that can be exploited by an attacker to perform unauthorized actions within the system. Information about known vulnerabilities can be found with a simple search of the Internet, as well as on such websites as the National Vulnerability Database (NVD [2015](#)). Data centers without effective patching processes risk exploitation through unpatched, known vulnerabilities by internal or external threats. The security community recognizes that zero-day vulnerabilities may not be disclosed to vendors, but sold to the highest bidder (e.g., data center attacker or a government agency) for future battles ([Harris 2014](#)). At the very least, known data center vulnerabilities should be patched.

From a defense perspective, a vulnerability scanner, such as the widely used Nessus scanner ([Nessus 2015](#)), provides particularly useful vulnerability data. The most useful scan provided by a vulnerability scanner is a credentialed scan, in

which the scanner logs into a system with sufficient privileges to perform a security assessment and determines whether the system has outstanding vulnerabilities that would not be otherwise detected by a noncredentialed scan.

Network Flow Data

The primary attack vectors for the data center are Internet-based and can be analyzed with network flow analysis using a tool such as Argus (2015). The network flows may contain information on successful and unsuccessful logon attempts or evidence related to malicious scans initiated by “bad actors,” which create and distribute malware and attempt to gain intelligence about the state of the data center. The flows could also contain command and control (C&C) traffic generated by malicious code that has implanted itself in the organization that needs to communicate with a command center. This malicious code could be an advanced persistent threat (APT), a threat that utilizes a sophisticated process in which attack code is implanted in a data center to perform long-term exfiltration of information and other malicious acts initiated by its remote C&C center.

This entry focuses on those security events for which the data center’s security team would be expected to take some manual action or initiate an automated response to address the threat posed. These are called *actionable security events*. For the purpose of identifying actionable security events, the following network data can be useful for the security team’s analysis:

- Network flow data that provides the following for the remote systems interacting with the data center:
 - IP address for remote system and data center system
 - Port used for the remote system and data center system connection
 - Protocol used for the remote system and data center system connection
 - Organization and country from which the network access originated, using information based on the IP address
 - Other flow-related data such as flow state, flags seen, packets sent and received, bytes sent and received, and duration

The ability to associate a data center user with a network flow is very useful (as will be discussed in later sections) since it allows a user to be associated with not only a flow but also with the IP address and organization from which the user logged in.

DNS Data

DNS traffic inspection provides a means to identify suspicious and malicious activity that may not be detected in network flow data, intrusion detection system (IDS) events, and/or audit logs. While an IDS may detect domain name system (DNS) cache poisoning and amplification attacks, the suspicious and malicious activities associated with DNS fast-flux, domain-flux, domains generated by domain-generation algorithms, and known malicious domains are detected by DNS traffic analysis. More will be said about this later in the entry.

Intrusion Detection System Data

An IDS, such as Snort (2020), provides a means to identify intrusion events based on patterns in the network flow. These patterns are constantly updated to detect new attacks. For an organization with a homogeneous set of systems, such as a Microsoft Windows shop, the IDS can be useful since only those patterns associated with Windows attacks can be loaded into the IDS. For an organization such as the data center that contains a heterogeneous set of systems, the IDS can be a two-edged sword in the battle to detect actionable events; while the IDS can identify relevant attack information, it can also generate a large amount of false-positive alerts that may be the result of an IDS pattern showing up in an encrypted stream of protected data that is flowing into and out of the data center. IDS can be useful, however, if properly used, as will be discussed later in this entry.

Centralized Log Data

Centralized log servers provide a valuable asset to the data center since they consolidate the log data from the plethora of systems that are used by the data center. Log data provides valuable information on who is doing what on the system, including who is performing remote logins to the data center. Also, of great importance, as will be shown

later, this login data allows an IP address to be associated with an authenticated user.

Risk-Partitioning the Mountain of Data

An important first step toward identifying the actionable data that may be hidden in the data associated with the operation of the data center is to partition the data into groups based on their potential for containing an indication of an actionable security event. Flow data represents the greatest volume of data, and so it will be the target for data reduction. This data will be partitioned into the following three categories:

- **Acceptable Risk Group:** Where there is an acceptable risk that the data does not contain any data that could indicate an actionable security event. While it is possible that some actionable data might also be contained in this group, the goal is to reduce this possibility to an acceptable level of risk.
- **High Risk Group:** Where there is a high risk that the data contains actionable data.
- **Unknown Risk Group:** Where there is no initial basis for assigning this flow data to either of the other two groups.

The following sections will explain an approach to perform this risk partitioning of flow data.

Partitioning Based on Acceptable Risk

Because network flows into a data center originate from IP addresses external to the data center, flows of acceptable risk can be partitioned from the rest by identifying acceptable-risk IP addresses. A subset of acceptable-risk IP addresses could be identified based on those from which a legitimate, authenticated user logs in to the data center. Furthermore, the acceptable risk attributed to the authenticated individual user could be extended to entire organization associated with IP address, provided it is a major national or international organization. Using readily available geo-location information, the organization associated with this IP address can be easily determined. Additionally, reusable IP addresses,

such as dynamically assigned Dynamic Host Configuration Protocol (DHCP) IP addresses, would be covered by this approach.

An exception to categorizing an entire organization as an acceptable risk based on one authenticated user would be made when the association between an authenticated user and an organization is transitory. A user logging into the data center from organizations with publicly accessible IP addresses, such as a public library, coffee shop, or hotel, should not result in these organizations being placed in an acceptable risk category. This exception will require manual screening against transitory affiliations.

Any network flow that falls into the acceptable risk category will be eliminated from intensive monitoring, further reducing the volume of the data that must undergo extensive scrutiny to identify actionable data.

Partitioning Based on High Risk

The next stage of the data risk-partitioning process is to focus on the data associated with high-risk sources. These are IP addresses and organizations that appear on watch lists of hostile sites, such as a malware domain list or a command and control (C&C) list, provided by external intelligence sources. These lists would include any organization that was known to have performed a network attack against a friendly government, university, or company. If an entire organization is known to be hostile, then all of the IP addresses associated with that organization are considered to be high risk.

Another potential indicator of a high-risk IP address or organization is the scanning of “dark space,” that is, unallocated IP addresses within the data center’s own assigned IP address space and/or transport protocol numbers on a host within the data center with no assigned service. A scan of a data center’s dark space originating from an external IP address indicates at least three possible scenarios:

- An error
- A legitimate mapping to the data center site
- A reconnaissance by a hostile agent

In general, scans that hit only a single or a small number of the data center's dark IP addresses or ports are assumed to be errors (the first scenario), and the IP addresses from which these scans originate are therefore classified as acceptable risk.

The problem of risk classification then lies in differentiating between the two remaining scenarios: a legitimate mapping of the data center's IP address space and a hostile reconnaissance of the space. Since scans in these two cases tend to look the same in that a large number of dark IP addresses or ports are hit, the conservative approach is to initially classify the originating IP address and hosting organizations as hostile and high risk. To minimize mis-categorization, a human analyst would periodically move any organizations that scanned dark space for legitimate network research purposes to a white list of approved scanners. An organization could also automatically be moved to the white list if an authenticated user logs in from this organization. Once an organization is on the white list, then for a specified period (e.g., 3–6 months) it would not be automatically reverted to the dark-space hostile list in response to future scans. At the end of the specified period, a determination should be made as to whether the organization will remain on the white list. These approaches are consistent with a “practical” situational awareness system.

Partitioning Based on Unknown Risk

Once the dark space analysis has been completed, the third stage of the risk-partitioning process involves the removal from the data all the flows that are comprised only of a scan, since scans alone are not considered an actionable event.

Results of Partitioning

The end result of risk partitioning is the elimination of the following data center flows:

- All flows that are nothing more than scans
- All flows from acceptable risk organizations

Following the removal of the above flows, the remaining flows would then contain only the following:

- All flows categorized as originating from high-risk organizations
- All flows that fall neither into the acceptable risk category or the high-risk category – unknown risk

The scans originating from IP addresses and organizations in the high-risk category will undergo significant scrutiny to determine whether they contain any actionable security events.

Once the re-categorization has gone as far as possible, the unknown risk flows will be subject to the same extensive scrutiny already given to the high-risk network flows. An important goal of risk partitioning is to move organizations and their associated flows from the unknown risk category into either the acceptable risk or the high-risk category, using any additional knowledge gained through scrutiny of their actions. Such scrutiny could reveal that a given organization is a supplier to or otherwise maintains a business relation with the data center, or that it is related in some meaningful way to a data center customer. Determinations of this sort will require manual analysis.

Whereas the current section addressed methods to filter and reduce the total volume of data, the next sections consider the analysis that can be used to identify actionable security events within the remaining data.

Data Enrichment to Facilitate Actionable Security Event Detection

The security-relevant data examined by the SSAS can be enriched through pre-processing to facilitate the detection of actionable security events. The first preprocessing step is to structure the data into a form that can be readily ingested by the SSAS, identifying key fields to achieve a uniform format. This uniform format, with extensions for particular types of data, will accommodate most, if not all, of the data to be used by the SSAS.

Splunk (2020) provides a powerful search engine on which to build a portion of the SSAS. Examples of useful fields for a Splunk-based SSAS project, shown in Fig. 1, describe network

Flow	IDS	Syslog
alertName	alertName	alertName
alertStatus	alertStatus	alertStatus
date_hour_min_sec	date_hour_min_sec	date_hour_min_sec
dstAddr	dstAddr	dstAddr
dstAsset	dstAsset	dstAsset
dstCategory	dstCategory	dstCategory
dstCity	dstCity	dstCity
dstCntry	dstCntry	dstCntry
dstNetwork	dstNetwork	dstNetwork
dstOpSys	dstOpSys	dstOpSys
dstOrganization	dstOrganization	dstOrganization
dstPort	dstPort	dstPort
dstRisk	dstRisk	dstRisk
dstVulnID	dstVulnID	dstVulnID
dstVulnSeverity	dstVulnSeverity	dstVulnSeverity
host	host	host
ipProto	ipProto	ipProto
riskDesc	riskDesc	riskDesc
source	source	source
sourcetype	sourcetype	sourcetype
srcAddr	srcAddr	srcAddr
srcAsset	srcAsset	srcAsset
srcCategory	srcCategory	srcCategory
srcCity	srcCity	srcCity
srcCntry	srcCntry	srcCntry
srcOpSys	srcOpSys	srcOpSys
srcOrganization	srcOrganization	srcOrganization
srcPort	srcPort	srcPort
srcRisk	srcRisk	srcRisk
srcVulnID	srcVulnID	srcVulnID
srcVulnSeverity	srcVulnSeverity	srcVulnSeverity
tags	tags	tags
flowDstBuffer	flowDstBuffer	
flowDstBytes	flowDstBytes	
flowDstPackets	flowDstPackets	
flowDstState	flowDstState	
flowIpFlags	flowIpFlags	
flowSessionID	flowSessionID	
flowSrcBuffer	flowSrcBuffer	

Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment,
Fig. 1 SSAS schema data field summary

flowSrcBytes	flowSrcBytes	
flowSrcPackets	flowSrcPackets	
flowSrcState	flowSrcState	
flowTcpOptions	flowTcpOptions	
flowDirection		
flowDuration		
	idsClassification	
	idsPriority	
	idsRuleGid	
	idsRuleName	
	idsRuleRevision	
	idsRuleSid	
	timeDiff	
		syslogCMD
		syslogDstUser
		syslogMessage
		syslogPID
		syslogProcess
		syslogSession
		syslogSrcUser
		syslogTTY

Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment, Fig. 1
(continued)

flows, IDS, and system log data. Some of these fields are common to all three of these data types, while others are more type specific.

After the data has been uniformly structured, it is augmented with context data to increase its relevance and meaning in relation to the environment (Shaw 2011). This introduces additional data points that can facilitate searching, grouping, and correlating event data, as well as writing alerts for specific events of interest.

Adding context to the data before it is ingested into the SSAS renders the data historically accurate, ensuring that the context for that data always reflects the state at ingestion, independent of subsequent changes. Context changes over time due to the evolution of the IT environment; for example, IP addresses are re-used, a system’s category may change, and vulnerabilities are eliminated as systems are patched.

Various types of context can be attached to event data, from geo-location to network access/location category (e.g., DMZ, external, internal, private, public) and by function (e.g., accounting, high-end computing, online storage, archival storage) with some indication of their importance.

Another important context is vulnerability. If a host has a vulnerability related to a service port, then any event data that references both that host and service port can be marked as suspicious, which allows the SSAS to use that context in determining if the event is actionable.

The next preprocessing step prior to ingestion is the identification of data associations among vulnerabilities, IDS events, and flows. For example, IDS data can enrich network flow data with IDS events based on source and destination attributes and time. Whereas a flow may be associated with multiple IDS events, an IDS event will be

associated with only one flow. Thus, identifying network flows that contain IDS events reduces to the simple process of searching the SSAS for network flows with IDS events. This shows one value of preprocessing, namely, that enriching the data before it is ingested by the SSAS reduces the work of the SSAS. Also of particular concern are flows that contain an IDS event that can exploit a target system vulnerability. By preprocessing the data to identify this association, unnecessary processing is avoided and performance of the SSAS is also enhanced.

Although the preprocessing described in this section will increase the overall volume of data, the goal of this preprocessing is to ease the compute load of the SSAS in identifying events of interest. Identifying events of interest is covered in the next section.

Detecting Actionable Security Events

In line with the goal of building a practical SSAS, we next look at the various types of actionable security events that should be detected from the data preserved from the partitioning described previously. The SSAS should be designed to detect actionable security states and events that are relevant to the enterprise being protected, such as the following:

- Network flows to ports with vulnerable services
- Intrusions
- Policy violations
- Insider threats and data exfiltration

Each of these will be discussed in the following subsections.

Vulnerabilities

An important step in securing data center systems is the remediation of known vulnerabilities that could allow a malicious agent to gain unauthorized access to or cause a denial of service on the systems. The vulnerabilities that need to be identified fall into the following broad classes:

- Vulnerabilities associated with the system's software
- Vulnerabilities due to misconfiguration of computer systems or networks
- Vulnerabilities that violate the security architecture or policies of the organization

Each of these vulnerability classes will be considered in what follows.

System's Software Vulnerability

System vulnerabilities can be detected using a vulnerability scanner such as Nessus (2015). The most useful vulnerability information can be obtained if the vulnerability scanners perform credentialed scans, wherein the scanners log into the system to identify vulnerabilities and determine the patches necessary to close those vulnerabilities. If a credentialed scan is not possible due to performance or availability concerns (e.g., the scan might crash a critical system), the vulnerability scanner can be configured to perform external scans of the organization's systems, though gaining much less information than through credentialed scans.

Using information from a credentialed scan of the data center's systems, the SSAS can produce information on the outstanding (unpatched) vulnerabilities associated with each system, including the risk that the vulnerability poses to the organization, as well as the time that has passed since a patch has existed for the vulnerability, but not applied. It can also identify other patches that can be applied to mitigate the vulnerability. Figure 2 shows a (static) spreadsheet from the SSAS that lists the vulnerabilities for several systems, categorized by the Nessus-assigned risk to the data center, which runs from Critical (the most severe risk), through High and Medium, to Low.

An embedded comment in the spreadsheet (not shown in Fig. 2) for the secure socket layer (SSL) vulnerability for the saturn.xxx.yyy reads: "The X.509 certificate chain for this service is not signed by a recognized certificate authority. If the remote host is a public host in production, this nullifies the use of SSL, as anyone could

Scan Date	Host	OS Vendor	Patch/Vulnerability	Risk	CVSS Score	Age in Days	Solution	CVE	CVE
Critical Patches/Vulnerabilities for CentOS Hosts									
Oct 13, 2014	mars.xx.x.yyy	CentOS	CentOS 5 / 6 / 7 : firefox / xulrunner (CESA-2014:1144)	Critical	10	41	Update the affected firefox and/or xulrunner packages.	CVE-2014-1562	CVE-2014-1567
High Patches/Vulnerabilities for CentOS Hosts									
Oct 13, 2014	venus.x.xx.yyy	CentOS	CentOS 5 / 6 / 7 : nss (CESA-2014:1307)	High	7.5	14	Update the affected nss packages.	CVE-2014-1568	
Oct 13, 2014	mars.xx.x.yyy	CentOS	CentOS 5 / 6 / 7 : procmail (CESA-2014:1172)	High	7.5	34	Update the affected procmail package.	CVE-2014-3618	
Medium Patches/Vulnerabilities for CentOS Hosts									
Oct 13, 2014	mars.xx.x.yyy	CentOS	CentOS 6 / 7 : openssl (CESA-2014:1052)	Medium	6.8	62	Update the affected openssl packages.	CVE-2014-3505	CVE-2014-3506
Oct 13, 2014	saturn.x.xx.yyy	CentOS	SSL Self-Signed Certificate	Medium	6.4	N/A	Purchase or generate a proper certificate for this service.		
Low Patches/Vulnerabilities for CentOS Hosts									
Oct 13, 2014	moon.x.xx.yyy	CentOS	CentOS 5 / 6 : samba / samba3x (CESA-2014:0866)	Low	3.3	97	Update the affected samba and/or samba3x packages.	CVE-2014-0244	CVE-2014-3493

Security Situational Awareness: A Practical Approach for a Complex Heterogeneous Environment, Fig. 2 Patches/vulnerabilities for CentOS hosts by risk

establish a man-in-the-middle attack against the remote host.” Also, the common vulnerability and exposures (CVE) entry in the source spreadsheet is linked to the CVE description in the National Vulnerability Database. For example, “CVE-2014-3505” for mars.xxx.yyy links to the associated vulnerability summary at: <https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2014-3505>.

Vulnerability information obtained by the scan can facilitate patch management. The number and age vulnerabilities should be monitored, since they could indicate changes that render a system susceptible to a successful attack. Thus, if the SSAS reveals that the average age of vulnerabilities is growing, problems with the data center’s patch management may exist. A sudden increase in the number of vulnerabilities could indicate the addition of systems with old, unpatched software.

Vulnerabilities Due to Misconfiguration of Computer Systems or Networks

Misconfiguration of the data center’s computer systems, for example, the failure to periodically change the default administrator user ID and password, could increase the risk of unauthorized access. A misconfiguration of the network access control lists (ACLs) on network routers could allow unwanted network traffic into the system, violating the organization’s network security policies. The SSAS could easily check for network misconfiguration by monitoring the network flow for ACL violations such as flows to low-order ports from the Internet. In addition, session-aware, “stateful” firewalls (any firewall that performs stateful packet inspection (SPI) or stateful inspection) can be used to eliminate flows that violate organizational policies or are unsolicited.

Vulnerability-Related Policy Violation

The SSAS can be used to detect a violation of the organization's security policy, such as the addition of a new system to the data center's production network that has not undergone the data center's configuration and approval process. The SSAS can determine a security compliance of a newly attached system by consulting a list of approved systems.

Intrusions

In this section, the ways in which the SSAS can identify events indicating an intrusion or detect strong evidence of a precursor to a persistent attempt to gain unauthorized access are discussed. Various types of intrusion events will be considered in the following subsections.

Intrusion Detection System Events

In a data center that supports a wide variety of operating systems for different customers, the use of the traditional IDS becomes problematic due to the plethora of IDS rules needed to support multiple OSes. Reducing the number of IDS rules to the level typical in a homogeneous computing environment, such as one comprised exclusively of Windows machines where only Windows IDS rules would be used, is nontrivial. Data centers with heterogeneous systems can manifest a multitude of IDS rules that misfire, increasing the difficulty in distinguishing actionable events from non-actionable events or false positives. However, because the data reduction described previously mitigates this issue, only the IDS events associated with high risk and unknown risk flows are considered here. In particular, the IDS-network flow data association can be used for reducing false positives, since a Windows IDS rule that fired on a flow destined for a Linux machine could be ignored. Those IDS events associated with high-risk flows are the highest priority for analysis, but the IDS events associated with unknown flows also require extensive scrutiny.

Brute-Force Attacks

Most systems connected to the Internet experience a constant barrage of attacks, starting with

scans and followed by login attempts. As noted above, scans are removed as part of the data reduction, leaving the SSAS to focus on the log analysis to detect brute force attacks. If the data center uses two-factor authentication, such as approaches that require both a personal identification number (PIN, something you know) and a hardware token (something you have) that produces a temporary password, then the risk of intrusion from a brute-force logon attempt is reduced. An agent attempting to penetrate the data center would have to obtain not only the PIN but also the token.

Suspicious Logins

The context (e.g., location of the system) of data, as previously described, is vital to the SSAS in flagging suspicious logins. Examples of a suspicious login include a login from a source IP address never seen before, a login from a country that may raise concern, or a login by a user from the mailroom into a finance server.

SSAS detection of a login from a source never seen before can be handled by recording the login site, but unfortunately, each user login from a new IP address will be initially regarded as a suspicious login. Using geo-location information that has been added to the data, the SSAS can determine if a new login comes from the same non-transitory organization from which a user has already logged in, which would not raise any suspicion. However, a login from multiple organizations by the same user account might warrant ongoing attention, especially if the logins come from widely separated geographic locations over a short period of time.

The use of geo-location information allows the SSAS to detect logins from countries that might be of concern if the SSAS has been provided with a list of countries of concern.

To address the third example, identifying a login from the mailroom to a financial system would require that the proper context for network category, system function, and user access be added to the data so that the SSAS, using a policy or set of rules, could identify this unwanted activity.

APTs, Malware Beacons, and Back Doors

After infecting a data center, advanced persistent threats (APTs) and malware beacons make periodic or random attempts to contact command and control (C&C) centers.

The SSAS applies time-trending analysis to data flow connections from the data center to remote systems in search of repeated contacts with the same source and destination IP addresses at unknown or high-risk organizations, which may indicate APTs and beacons. Due to the data-reduction risk-partitioning process described in the previous section, the flows requiring assessment are greatly reduced, leaving only high-risk and unknown data flow.

While convenient for special access to otherwise secure systems under exceptional circumstances (e.g., for law-enforcement), backdoors allow an intruder to gain and maintain ongoing unauthorized access to these systems if compromised. Such backdoors could be used by a C&C center to contact APTs on a host, providing it with instructions. One approach to defeating compromised backdoors is to enforce “deny all, allow specific” ACL or firewall rules and monitor the allowed inbound connections that do not correspond to any known services.

If a user’s account has been compromised, then the beaconing activity would not be discovered under the approach describe here, since flows from the account would be categorized as accepted risk and eliminated from assessment during risk-partitioning. However, the goal is to develop a practical approach to intrusion detection. Provided the data center uses two-factor authentication and keeps its patching up to date, then beaconing from a compromised account has a fairly low probability of occurring. If such beaconing is a concern, then Hadoop could be unleashed on accepted risk data center connections to look for suspicious activity.

Suspicious DNS Activity

An effective SSAS should detect malicious and suspicious DNS activity such as the use of known malicious domains, “fast-flux,” “domain-flux,” and domain name generation via domain generation algorithms (DGA).

The use of known malicious domains is by far the easiest to detect. Comparing the DNS queries from data center systems against a list of known malicious domains could expose malicious code, such as an APT.

Detecting fast-flux and domain-flux activity is more difficult. Fast-flux is a malware cloaking technique that hides malware delivery sites and C&C servers behind a common legitimate domain name that is associated with hundreds or even thousands of IP addresses. These IP addresses are changed at an extremely high rate using a short time-to-live (TTL) value in the DNS resource record. DNS queries that return a large number of IP addresses for the same domain name may indicate the presence of malware. The threshold for signaling an abnormal number of IP addresses for a given host should be higher than the number for such large sites as Google, Facebook, and other popular content delivery networks.

Domain-flux is a malware cloaking technique that uses a large volume of randomly generated domain names where only a few resolve to an actual IP address, masking malware delivery sites and C&C servers. An unusually large number of failed domain name lookups originating from the data center system over a period of time may indicate the presence of malware.

Detecting domain names generated by DGAs can be difficult due to the random characters in the names. While some DGA-generated names may be a series of numbers or a combination of two or more dictionary words, other DGA-generated names appear as unintelligible strings of characters. Using machine learning, one could create a detection algorithm that would flag domain names likely generated by DGAs in DNS queries within the data center, and thus, detect possible malware.

Policy Violations

The SSAS can also be used to detect security policy violations. For example, to ensure individual accountability, a security policy may exist in which system administrators in a data center are to be held to auditable, individually accountable actions, that is, actions that can be traced to a particular individual. If a system administrator logs directly into a system as root to perform privileged actions, this policy is effectively

violated because there is no individual accountability due to the shared use of the root account among a number of system administrators. However, if the system administrator first logs onto the system with his or her individual account and then uses sudo to perform the privileged action, then individual accountability is preserved.

The SSAS could detect when this policy is violated via the system logs that would show when system administrators logged in directly as root. Although the logs would indicate a definite violation of policy, they would not directly identify the system administrator who violated the policy. However, indirect log information, such as the login time, can be used to identify the origination sites of network flows that were used for the root login, which can be compared with previously recorded login sites where the users are known. In this way, the set administrators who may have logged in directly with root in violation of the organization's security policy can be greatly narrowed.

The SSAS could also be used to address a security policy that prohibits the sharing of accounts on a data center system, which would include the unauthorized access or the unauthorized sharing of access credentials. While the SSAS will not be able to identify such sharing in all cases with high certainty, it can detect the possibility of such sharing in some cases, such as in the case of multiple logins into the same account from different IP addresses within a short period of time, especially for IP addresses with significant geographic separation. Unusual time differences in logins may also indicate account sharing, for example, logins in the middle of the business day, followed later very late at night at the primary location for the user.

Finally, the SSAS could be used to enforce a security policy requiring users to be actively using the systems in which they are logged in throughout the login session. Unattended logins are of concern from a security standpoint because a logged-in user's end-system could provide a malicious agent an entry point into an organization's data center system. In this case, the agent does not have to penetrate the organization's system itself, but only the remote system that is the entry point for a long-duration login. An example of an attack that could use such a connection is one in which a

malicious agent on the same remote system as a legitimate user exploits the SSH multiplexing capability to piggyback into the data center on the "back" of an existing connection. The SSAS can easily detect such long-duration sessions and notify security staff when these sessions exceed some specified limit that violates organizational policy.

Insider Threats and Data Exfiltration

The insider threat, including that leading to data exfiltration, can be difficult for a data center that supports many different groups and projects to address. Increased group and project diversity reduce the commonality in characteristic actions, the departure from which would normally indicate that an insider is misusing data center data or processing or is attempting to leak sensitive data. In such data centers, even external attacks against the data center can be a challenge to identify.

However, aberrant user behavior may indicate harmful actions such as system misuse or data exfiltration. This behavior includes unusual login times, the addition of a new source address for a remote user, successful or attempted access to an unauthorized host, and successful or attempted privilege escalation. One approach to detecting aberrant user behavior is to build user profiles, which include user login times and source addresses. Once built, these profiles can be compared against network flow data and audit records to identify events that indicate aberrant user behavior, which would generate an alert.

While aberrant user behavior may indicate data exfiltration due to insider threat or compromised accounts, unusual outbound activity may also indicate data exfiltration not only by an insider, but also by a compromised account or an APT. Indicators of unusual outbound activity may be:

- Outbound connections to destinations not associated with the user
- Outbound connections with large volumes of data sent at unusual times or to destinations not associated with users
- Long-duration outbound connections that send a modest amount of data designed to go unnoticed
- Large volumes of outbound emails to free email providers

- Email bursts to a large number of accounts, where each email is unique
- Extremely large email attachments

Note that users can employ more covert data exfiltration methods by hiding sensitive data within conventional data.

Future Directions

Future directions for improving security situational awareness in complex heterogeneous environments include:

- The development of intelligent automation to replace human involvement in the current approach, while maintaining low rates of both false positive and false negative alerts
- The refinement of approaches to partitioning data center network flow exclusively into actionable security events
- The development of new security monitoring capabilities that allow for the exchange of

threat information with other organizations that enhance the rapid identification and automated response to threats that will allow the security staff to move from a reactive to a more proactive approach to security

Bibliography

- Argus (2015) Audit record generation and utilization system. <http://qosient.com/argus/>
- Harris SS (2014) @War: the rise of the military-internet complex. Eamon Dolan/Houghton Mifflin Harcourt, New York
- Nessus (2015) Tenable Network Security. <http://www.tenable.com/products/nessus-vulnerability-scanner>
- NVD National Vulnerability Database (2015) National Institute of Standards and Technology. <https://nvd.nist.gov>
- Shaw D (2011) Reducing false-positives in security event data using context, Presentation at the NASA IT summit, San Francisco. [http://www.nasa.gov/ppt/583349main_2011_Present_NASA_IT_Summit_Shaw_Reducing_False_Positives_\(2\).ppt](http://www.nasa.gov/ppt/583349main_2011_Present_NASA_IT_Summit_Shaw_Reducing_False_Positives_(2).ppt)
- Snort (2020). <https://www.snort.org>
- Splunk (2020). <http://www.splunk.com>

Security with Privacy

Elisa Bertino
CS Department, Purdue University, West
Lafayette, IN, USA

Article Outline

Introduction
Privacy Enhancing Techniques for Security
Future Directions
Concluding Remarks
Glossary
References

Introduction

Recent technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, and social networks, are making possible to capture, process, and share huge amounts of data – referred to as *big data* – and to extract useful information from this data and predict trends and events (Bertino 2013). Big data in particular is critical for security-related tasks (Bertino 2014). For example, by analyzing and integrating data collected in social networks one can identify connections and relationships among individuals that may in turn help with homeland protection. By collecting and mining data concerning user travels and disease outbreaks one can predict disease spreading across geographical areas. In cyber security, big data technologies can help with system and network security monitoring, situation awareness, anomaly detection, user monitoring, protection from insider threat (Bertino 2012), continuous user authentication, and context-based access control.

And those are just a few examples; many other domains exist in which big data technologies can play a major role in enhancing security.

The use of data for security tasks, however, raises major privacy concerns. Collected data, even if anonymized by removing identifiers such as names or social security numbers and by using generalization techniques (Dai et al. 2009), when linked with other data may lead to re-identify the individuals to which specific data items are related to. Also, as organizations, such as governmental agencies, often need to collaborate on security tasks, data sets are exchanged across different organizations, resulting in these data sets being available to many different parties, thus increasing the data exposure to privacy breaches. Security tasks such as authentication and access control may require detailed information about users. An example is multifactor authentication that may require, in addition to a password or a certificate, user biometrics. Recently proposed continuous authentication techniques extend user authentication to include information such as user keystroke dynamics in order to constantly verify user identity. Another example is location-based access control (Damiani et al. 2007) that requires users to provide the access control system information about their current location. As a result, detailed user mobility information may be collected over time by the access control system. This information if misused or stolen can lead to privacy breaches.

It would thus seem that privacy and security are conflicting requirements. If we want to achieve security, we must give up privacy; on the other hand, if we are keen on assuring privacy, we may undermine security. However, this may not be necessarily the case. Recent advances in applied cryptography are making possible to work on encrypted data – for example, for performing analytics on encrypted data (Liu et al. 2014; Yi et al. 2015). However, much more needs to be done as the data privacy techniques to use heavily depend on the specific use of data and the security tasks at hand. Also current techniques are not still

The presentation in this entry is partially based on the panel statement paper [2]

© Springer Science+Business Media, LLC, part of Springer Nature 2023
T.-Y. Lin et al. (eds.), *Granular, Fuzzy, and Soft Computing*,
https://doi.org/10.1007/978-1-0716-2628-3_754

Originally published in
R. A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*, © Springer Science+Business Media LLC 2021
https://doi.org/10.1007/978-3-642-27737-5_754-1

able to meet the efficiency requirement of many applications.

In this entry, we elaborate on whether security and privacy can be reconciled. We first present a few examples of privacy-enhancing techniques that help in reconciling security and privacy. We then discuss relevant future directions for security with privacy.

Privacy Enhancing Techniques for Security

Many privacy enhancing techniques have been proposed over the last 20 years, ranging from cryptographic techniques, such as oblivious data structures (Wang et al. 2014) that hide data access patterns, to data anonymization techniques that transform the data to make more difficult to link specific data records to specific individuals (Byun et al. 2007). The problem of location privacy has also been the focus of extensive research both in the past and presently (Ghinita et al. n.d.; Paulet et al. 2014; Damiani et al. 2010). More recently research efforts have been devoted to investigate privacy-preserving techniques for data on the cloud (Nabeel and Bertino 2014a; Seo et al. 2014), on smart phones (Shebaro et al. 2014), and on social networks (Carminati et al. 2013). We refer the reader to specialized conferences, such as the Privacy-Enhancing Symposium (PET) (<https://petsymposium.org/2014/>) series, and journals, such as *Transactions on Data Privacy* (<http://www.tdp.cat/>) and *IEEE Transactions on Dependable and Secure Computing* (<http://www.computer.org/web/tdsc>), for further references. However, it is important to note that most proposed privacy-enhancing techniques only focus on privacy and do not address the key problem of reconciling security with privacy. A few notable exceptions and in what follows we focus on some of these approaches.

Privacy-Preserving Data Matching

Record matching is typically performed across different data sources with the aim of identifying common information shared among these sources. An example is matching a list of passengers on a

flight with a list of suspicious individuals. However, record matching from different data sources is often in contrast with privacy requirements concerning the records to be matched. Cryptographic approaches, like secure set intersection protocols (Freedman et al. 2004), may alleviate such concerns. Secure set intersection protocols allow two parties, each owning a set, to privately compute the intersection of these sets without requiring each party to disclose to the other party its own set. The only information that is shared between the two parties is the set of elements in the intersection. Many protocols have been developed for computing secure set intersection. However, those protocols do not scale for large data sets. Experimental results from (Rao et al. 2019) show that on a single Intel Core i7-2600 CPU machine it takes 974.4 h to execute the private linkage of two datasets, each having 6,000 records and 5 attributes.

A recent matching protocol (Rao et al. 2019) combining secure multiparty computation (SMC) techniques and differential privacy (Dwork and Roth 2014) addresses scalability issues and also security issues of previous protocols. Such novel protocol is based on a hybrid strategy by which each party partitions its own dataset into subsets and releases a synopsis of each such subset (i.e., the subset extent and size) to the other party. Record matching between “faraway” subsets is then pruned, and thus efficiency is greatly improved. To protect the privacy of the records in each subset, the subset size is randomized by differential privacy (Dwork and Roth 2014) before being released, which ensures that the presence or absence of any individual record does not affect the released subset size significantly. More specifically, the matching protocol includes three parties: Alice and Bob are the data set owners, and Charlie is a third party. The three parties are *honest-but-curious* (Goldreich 2004). Alice and Bob partition their datasets into subsets and send Charlie the differentially private synopses of the subsets. Charlie coordinates the record matching between Alice and Bob. Based on the received synopses, Charlie prunes the record matching between subsets with a distance beyond a threshold specified by Alice and Bob. During the entire

protocol, Charlie does not access any record value. At the end of the protocol, its knowledge remains limited to the private synopses of the subsets released to it by Alice and Bob. At the same time, Alice and Bob obtain the matching records. But they neither learn the attribute values of non-matching records nor the private synopses of the subsets. Such hybrid protocol thus separates the differentially private synopses from the matching records. The protocol has been formally proved to comply with differential privacy. The protocol also uses a private indexing scheme to enhance the effectiveness of data partitioning. The indexing scheme hierarchically partitions a dataset, forming a tree. Whenever a node in the tree needs to be split, a privacy budget is dynamically allocated based on the node size, in such a way that the magnitude of added noise does not dominate the node size. In other words, the indexing scheme optimizes the accuracy of the noisy node size in relative terms. Experimental results show that such hybrid protocol is superior to existing approaches with respect to both recall and efficiency in the context of private record linkage. The execution of the hybrid protocol for computing the privacy-preserving record matching for two datasets, each having 6,000 records and 5 attributes, takes 20.17 h which is an improvement of 2 orders of magnitude compared with the 974.4 h required with a conventional private matching protocol.

However, even though such approach represents an important step, research is needed along several directions. Privacy-preserving techniques must be designed that are suited for complex matching techniques, based for example on semantic matching. Security models and definitions also need to be developed supporting security analysis and proofs for solutions combining different security techniques, such as SMC and differential privacy.

Privacy-Preserving Collaborative Data Mining

Conventional data mining is typically performed on big centralized data warehouses collecting all the data of interest. However, centrally collecting all data poses several privacy and confidentiality concerns when data belongs to different

organizations. An approach to address such concerns is based on distributed collaborative approaches by which the organizations retain their own data sets and cooperate to learn the global data mining results without revealing the data in their own individual data sets.

Fundamental work in this area includes: (i) techniques allowing two parties to build a decision tree without learning anything about each other data sets except for what can be learned by the final decision tree (Lindell and Pinkas 2000); (ii) specialized collaborative privacy-preserving techniques for association rules, clustering, k-nearest neighbor classification (Vaidya et al. 2006). Usually most of the approaches proposed for privacy-preserving collaborative data mining are based on cryptographic tools such as oblivious transfer protocols (Even et al. 1985; Naor and Pinkas 2001), oblivious evaluation of protocols (Naor and Pinkas 2006), and oblivious circuit evaluation (Yao 1986).

These tools are however still very inefficient. Recent research has thus been focusing on more efficient implementation techniques for basic cryptographic tools. Notable recent examples include approaches to enhance the efficiency and scalability of secure function computation by circuit evaluation (Kreuter et al. 2013; Demmler et al. 2015). However, research is needed on how to efficiently use such novel implementations in the context of data mining. Novel approaches based on cloud computing and new cryptographic primitives also have to be investigated. Notable recent work includes an efficient approach to perform clustering on data stored in a cloud (Liu et al. 2014). This approach however needs to be extended for use in a collaborative setting.

Privacy-Preserving Biometrics Authentication

Conventional approaches to biometrics authentication require recording biometrics templates of enrolled users and then using these templates for matching with the templates provided by users at authentication time (Jain and Ross n.d.). Templates of user biometrics represent sensitive information that needs to be strongly protected. In distributed environments in which users have to interact with many different service providers the

protection of biometric templates becomes even more complex. A main problem is how to support strong authentication via biometrics without revealing in clear the biometric template to the identity verifier. Notice that in a distributed system supporting decentralized authentication multiple verifiers may have to have available such templates.

A recent approach addresses such issue by using a combination of perceptual hashing techniques, classification techniques, and zero-knowledge proof of knowledge (ZKPK) protocols (PrivBioMTAuth 2018). Under such approach, the image of the biometrics, for example the image of a retina scan, of the user is processed to extract from it a string of bits through the use of a perceptual hash function (Klinger and Starkweather n.d.) based on the Discrete Cosine Transform. Such string of bit is then processed through a support vector machine (SVM) classifier. The label of the class returned by the SVM classifier is then combined with a secret S derived from a user's password. The result of such combination is the biometric identifier (BID) of the user. The class label is an integer represented by 32-bits and S is 128-bits. The generated BID (which is 160-bits) is then used together with a random number to generate a Pedersen cryptographic commitment (Pedersen 1992) which represents an identification token that does not reveal anything about the original input biometrics. The commitment is then used in the ZKPK protocol to authenticate the user. To enhance accuracy error Hadamard error correcting codes are also used. The approach has been engineered for secure use on mobile phones. Experimental results show that the accuracy of the proposed biometric-based authentication is at 90% when 16-bit Hadamard codes are used.

Much work needs however to be done to increase accuracy. Different approaches to authentication need to be investigated based on recent homomorphic encryption techniques. Also, recent trends in authentication suggest the use of multimodal biometric authentication, by which multiple biometrics are used to authenticate the users, and continuous authentication, by which the system learns features from the behavior of

the users and then compares the actual user behavior with the learned features. An example of such feature is the speed with which the user types at a computer board. However, those advanced forms of authentication further increases the need of collecting detailed data about each user. Privacy-preserving techniques need to be devised supporting such forms of authentication, such as techniques similar to the technique described in this subsection or based on recent homomorphic encryption techniques.

Privacy-Preserving Fine-Grained Access Control for Cloud Data

Cloud technology allows a large variety of organization to store and manage large amount of data at lowest costs than was possible before. However, confidentiality of data stored in a cloud is a critical requirement, especially when the cloud is public and manages data and applications by different customers. Encryption is a commonly adopted approach to protect the confidentiality of the data. Encryption alone, however, is not sufficient as organizations often have to enforce fine-grained access control on the data. Fine-grained access control is one of the most critical security functions. Such control is often based on attributes of data users, referred to as identity attributes, such as the roles of users in the organization, projects on which users are working and so forth. These access control systems are usually referred to as attribute-based access control (ABAC) systems. The well-known XACML standard (XACML n.d.) is a well-known example of an ABAC model. We refer the reader to (Bertino et al. 2011) for a survey of access control, including the ABAC model. Therefore, an important requirement is to support fine-grained access control, based on ABAC policies, over encrypted data stored on the cloud.

A major issue in addressing such requirement is that the identity attributes used in the ABAC policies often reveal privacy-sensitive information about users and leak confidential information about the data content. The confidentiality of the content and the privacy of the users are, thus, not fully protected if the identity attributes are not protected. Further, the privacy, both individual as

well as organizational, is a key requirement in all solutions, including cloud services, for digital identity management. Further, as insider threats (Bertino 2012) are one of the major sources of data theft and privacy breaches, identity attributes must be strongly protected even from accesses within organizations. With the adoption of cloud technology the scope of insider threats is no longer limited to the organizational perimeter. Therefore, protecting the identity attributes of the users while enforcing attribute-based access control both within the organization as well as in the cloud is crucial.

An approach to address the above requirement is based on a fine-grained encryption of data combined with a novel efficient approach for attribute-based broadcast group key management (AB-BGKM) (Nabeel et al. 2013). Fine-grained encryption allows one to encrypt different portions of the data with different keys while at the same time minimizing the number of encryption keys required; which portions of the data are encrypted with which keys is based on ABAC policies (Nabeel et al. 2013). The AB-BGKM scheme (Nabeel and Bertino 2014b) allows one to share encryption keys among a group of users. Keys are shared based on user identity attributes, so that each user in a group of users receives only the encryption keys for the data portions he/she is authorized to access. The AB-BGKM scheme has several relevant properties. When the group changes, the rekeying operations do not affect the private information of existing group members and thus the AB-BGKM scheme eliminates the need of establishing private communication channels. The scheme provides the same advantage when the group membership conditions change. Furthermore, the group key derivation is very efficient as it only requires a simple vector inner product and/or polynomial interpolation. Additionally, the AB-BGKM scheme is resistant to collusion attacks. Multiple group members are unable to combine their private information in a useful way to derive a group key which they cannot derive individually. The AB-BGKM scheme is based on the access vector technique (Shang et al. n.d.) and Shamir's threshold scheme (Shamir 1979). The access control vector

technique allows one to hide a secret, that is, a data encryption key, in some publicly available information, referred to as access control vector. Only users with certain secrets can extract the information from this vector, which can thus be published on the cloud. Users receive these secrets by engaging in an oblivious envelop transfer protocol with the identity issuer. The oblivious envelop transfer protocol assures that a user is able to extract the secret from the envelop only if his/her identity attributes verify the conditions associated with the envelop. To summarize, under the approach in (Nabeel et al. 2013), fine-grained encrypted data is stored on the cloud together with access control vectors hiding the encryption keys. From the access control vectors the cloud cannot learn the keys hidden in the vectors nor any information about the user identity attribute. When a user wants to access some data items from the cloud, the user downloads the encrypted data items and the corresponding access vectors. The user then using his/her own secret extract the encryption keys from the vectors. Notice that if a user does not have the authorization to access the downloaded encrypted data items, the user will not be able to extract the encryption keys from the access control vectors, as the access control vectors are generated based only on the secrets held by the users authorized to access the data items. If policies change or a user is not any longer authorized to access certain data items, it is sufficient to re-encrypt the data with some new encryption keys and generate new access control vectors, and upload them to the clouds. No specific actions or communications are required with users. We refer the reader to Nabeel et al. (2013) and Nabeel and Bertino (2014b) for details and performance evaluation.

Even though the approach described in (Nabeel et al. 2013) is an effective and efficient approach to support fine-grained access control, based on identity attributes, while at the same time assuring privacy of identity attributes of users, several issues still need to be addressed. A first issue is related to the problem of hiding data access patterns. By looking at the data access patterns, it is possible to derive useful information (Islam et al. 2012) from the data. Therefore, it is

critical to devise effective approaches to hide such access patterns. In addition, it is critical to develop approaches able to support context- and content-based access control that can be enforced on a cloud without leaking information about context and content.

Future Directions

Despite initial solutions to the problem of security with privacy have been proposed – some of which have been outlined in the previous section, comprehensive solutions require addressing many research challenges especially when dealing with big data. In what follows, we outline relevant future directions.

- **Data Confidentiality:** Data confidentiality is a critical requirement for data privacy. Several data confidentiality techniques and mechanisms exist – the most notable being access control and encryption. Both have been widely investigated. However, with respect to access control systems for big data we need approaches for:
 - *Merging large numbers of access control policies.* In many cases, big data entails integrating data sets originating from multiple sources; these data sets may be associated with their own access control policies, referred to as “sticky policies,” and these policies must be enforced even when a data set is integrated with other data sets. Therefore policies need to be integrated and conflicts solved possibly by using some automated or semi-automated policy integration system. EXAM is an example of such a system (Lin et al. 2010). Policy integration and conflict resolution are, however, much more complex when dealing with privacy-aware access control models, such as PRBAC (Ni et al. 2010), as these models allow one to specify policies that include the purpose for which the access to a protected data item is allowed, obligations arising from the use of data, and special privacy-related conditions that must be met in order to access the data. Automatically integrating such type of policies and solving conflicts is a major challenge.
 - *Automatically administering authorizations for big data and in particular for granting permissions.* If fine-grained access control is required, manual administration on large data sets is not feasible. We need techniques by which authorization can be automatically granted, possibly based on the user digital identity, profile, and context, and on the data contents and metadata. An interesting first step toward the development of such techniques is presented in (Ni et al. 2009). However, more advanced approaches are needed able to deal with dynamically changing contexts and situations.
 - *Enforcing access control policies on heterogeneous multi-media data.* Content-based access control is an important type of access control by which authorizations are granted or denied based on the content of data. Content-based access control is critical when dealing with video surveillance applications which are important for security. As for privacy such videos have to be protected. Supporting content-based access control requires understanding the contents of protected and this is very challenging when dealing with multimedia large data sources.
 - *Automatically designing, evolving, and managing access control policies.* When dealing with dynamic environments where sources, users, and applications as well as the data usage are continuously changing, the ability to automatically design and evolve policies is critical to make sure that data is readily available for use while at the same time assuring data confidentiality. Environments and tools for managing policies are also crucial.
 - *Enforcing access control policies in big data stores.* Some of the recent big data systems allow its user’s to submit arbitrary jobs using programming languages such as Java. For example, in Hadoop, users can submit arbitrary MapReduce jobs written

in Java. This creates significant challenges to enforce fine grained access control efficiently for different users. Although there is some existing work (Ulusoy et al. [n.d.](#)) that tries to inject access control policies into submitted jobs, more research needs to be done on how to efficiently enforce such policies in recently developed big data stores, especially if access control policies are enforced though the use of fine-grained encryption.

- Data Privacy: a major issue arising from big data is that correlating many (big) data sets one can extract unanticipated information. Relevant issues and research directions that need to be investigated include:
 - *Techniques to control what is extracted and to check that data is used for the intended use.* Content-based access control is one such technique in that it allows one to return certain data to a given user based on the contents of the data (Bertino et al. [2011](#)). Content-based access control is typically supported in DBMS through the use of view mechanisms (Bertino and Haas [1988](#)) or query modifications. Supporting content-based access control stored by systems other than DBMS is much more difficult because of the difficulty of characterizing the conditions that the data contents must verify in order to be returned to a user. In DBMS such conditions are easily expressed as SQL queries. Research is needed to design techniques able to support content-based access control for a variety of data management systems. Another more difficult question to address is how to verify that data, returned to a user, is used for the intended purpose. An initial pioneering approach was proposed that associates with each data item a set of possible purposes, from an ontology of purposes, for which the data can be used (Byun et al. [2005](#)). When a user accesses some data items, the user indicates in the access request the purpose(s) for which the data items are being accessed. The query purposes are then matched against the purposes associated with the data items to verify that the access is according to the intended use of data. Such an approach needs to be complemented with techniques for automatically and securely identifying the data access purposes, instead of relying on indications given by users as part of their access requests.
 - *Support for both personal privacy and population privacy.* In the case of population privacy, it is important to understand what is extracted from the data as this may lead to discrimination. Also when dealing with security with privacy, it is important to understand the trade-off of personal privacy and collective security.
 - *Efficient and scalable privacy enhancing techniques.* Several such techniques have been developed over the years, including oblivious RAM, security multiparty computation, multi-input encryption, homomorphic encryption. However, they are not yet practically applicable to large data sets. We need to engineer these techniques, using for example parallelization, to fine tune their implementation and perhaps combine them with other techniques, such as differential privacy, like in the case of the record linkage protocols described in (Yi et al. [2015](#)). A possible further approach in this respect is to first use anonymized/sanitized data, and then depending on the specific situation to get specific non-anonymized data.
 - *Usability of data privacy policies.* Policies must be easily understood by users. We need tools for the average users and we need to understand user expectations in terms of privacy.
 - *Privacy implication on data quality.* Recent studies have shown that people lie especially in social networks because they are not sure that their privacy is preserved. This results in a decrease in data quality that then affects decisions and strategies based on these data.
 - *Risk models.* Different types of relationship of risks with big data can be identified: (a) big data can increase privacy risks;

(b) big data can reduce risks in many domains (e.g., national security). The development of models for these two types of risk is critical in order to identify suitable tradeoff and privacy-enhancing techniques to be used.

- *Data ownership.* The question about who is the owner of a piece of data is often a difficult question. It is perhaps better to replace this concept with the concept of stakeholder. Multiple stakeholders can be associated with each data item. The concept of stakeholder ties well with risks. Each stakeholder would have different (possibly conflicting) objectives and this can be modeled according to multiobjective optimization. In some cases, a stakeholder may not be aware of the others. For example, a user to whom a data pertains (and thus a stakeholder for the data) may not be aware that a law enforcement agency is using this data. Technology solutions need to be investigated to eliminate conflicts.
- *Data lifecycle framework.* A comprehensive approach to privacy for big data needs to be based on a systematic data lifecycle approach. Phases in the lifecycle need to be identified and their privacy requirements and implications need to be identified. Relevant phases include:

Data acquisition. We need mechanisms and tools to prevent devices from acquiring data about other individuals which relevant when devices like Google glasses are used. For example we need mechanisms able to automatically block devices from recording/acquiring data when in certain location or notify a user that recording devices are around. We also need techniques by which each recorded subject may express his/her preferences about the use of the data.

Data sharing. Users need to be informed about data sharing/transferred to other parties. Always informing users however is not always possible as sometimes information about data transfer and use is confidential to the

organization's missions. It is thus critical to devise legal guidelines on such issue based on which technical mechanisms can be based.

Concluding Remarks

This entry has discussed the difficult problem of reconciling security with privacy. The entry has shown initial examples of technologies that pave the way toward effective approaches to this problem. The entry has also identified a number of key challenges that arise when trying to use big data while at the same maximizing the use of the data and preserving data privacy and confidentiality. Addressing the above challenges require multidisciplinary research drawing from many different areas, including computer science and engineering, information systems, statistics, risk models, economics, social sciences, political sciences, human factors, and psychology. We believe that all these perspectives are needed to achieve effective solutions to the problem of privacy in the era of big data and of how reconcile security with privacy.

Acknowledgments The work reported here has been partially supported by the Purdue Cyber Center (Discovery Park), by NSF under awards CNS-1111512 and CNS-1016722, and by NIST under the project “Advancing Commercial Participation in the NSTIC Ecosystem.” The contents of Section 3 are partially based on the results of the privacy session (chaired by the author of the present entry) organized as part of the NSF Big Data Security and Privacy Workshop, September 16-17, 2014, <http://csi.utdallas.edu/events/NSF/NSF%20workshop%202014.htm>. I would like to thank the workshop Chair, Professor Bhavani Thuraisingham, the chair of the workshop security session, Professor Murat Kantarcioglu, and the workshop participants for the discussions during the workshop which lead to an initial draft agenda for privacy research for big data.

Glossary

Data Matching It is the task of finding records that refer to the same entity.

Secure Multiparty Computation It is a class of techniques allowing two or more parties to jointly compute a function over their inputs

while keeping these inputs private from each other.

Privacy Preserving Collaborative Data Mining It is a special case of secure multiparty computation where the functions to be computed are data mining functions.

Access Control It refers to controlling actions executed by subjects on protected resources to ensure that only properly authorized subjects can execute the requested actions.

Attribute-Based Access Control It is a special form of access control in which access control decisions are based on properties, that is, attributes, of subjects, protected resources, and context.

Authentication It is the process of verifying the identity of a subject.

Biometric-Based Authentication It is a form of authentication that relies on the unique biological characteristics of an individual to verify the identity of the subject.

References

- Bertino E (2012) Data protection from insider threats. Morgan&Claypool
- Bertino E (2013) Big data – opportunities and challenges. Panel Statement, IEEE COMPSAC
- Bertino E (2014) Security with privacy – opportunities and challenges. Panel Statement, IEEE COMPSAC
- Bertino E, Haas LM (1988) Views and security in distributed database management systems. In: Proceedings of the international conference on Extending Database Technology (EDBT'88), Venice, Italy, March 14–18, 1988. Springer. Lecture Notes in Computer Science
- Bertino E, Ghinita G, Kamra A (2011) Access control for databases: concepts and systems. Found Trends Database 3(1–2):1–148
- Byun JW, Bertino E, Li N (2005) Purpose based access control of complex data for privacy protection. In: Proceedings of 10th ACM symposium on Access Control Models and Technologies (SACMAT 2005)., Stockholm, Sweden, June 1–3
- Byun J-W, Kamra A, Bertino E, Li N (2007) Efficiently k-Anonymization Using Clustering Techniques. In: Proceedings of 12th international conference on Database Systems for Advanced Applications (DASFAA 2007), Bangkok, Thailand, April 9–12. LNCS, Springer
- Carminati B, Ferrari E, Viviani M (2013) Security and trust in online social networks. Morgan&Claypool
- Dai C, Ghinita G, Bertino E, Byun J-W, Bertino E (2009) TIAMAT: a tool for interactive analysis of microdata anonymization techniques. PVLDB 2(2):1618–1621
- Damiani ML, Bertino E, Silvestri C (2010) The PROBE framework for the personalized cloaking of private locations. Transact Data Priv 3(2):123–148
- Damiani M, Bertino E, Catania B, Perlasca P (2007) GEO-RBAC: a spatially aware RBAC. ACM Trans Inf Syst Secur 10(1)
- Demmler D, Schneider T, Zohner M (2015) ABY- a framework for efficient mixed-protocol secure two-party computation. In: Proceedings of the 2015 Network and Distributed System Security (NDSS) symposium, San Diego, California, February 8–11. (in print)
- Dwork C, Roth A (2014) The Algorithmic Foundations of Differential Privacy. Found Trends Theor Comput Sci 9(3-4):211–407
- Even S, Goldreich O, Lempel A (1985) A randomized protocol for signing contracts. Commun ACM 28: 637–647
- Freedman M, Nissim K, Pinkas B (2004) Efficient private matching and set intersection. In: Proceedings of Eurocrypt, vol LNCS 3027. Springer, pp 1–19
- Ghinita G, Kalnis P, Khoshgozaran A, Shahabi C, Tan KL (n.d.) Private queries in location based services: anonymizers are not necessary. In: Proceedings of 2008 ACM SIGMOD International Conference on Management of Data
- Goldreich O (2004) The foundations of cryptography – volume 2, Basic Applications. Cambridge University Press
- Islam MS, Kuzu M, Kantarcioglu M (2012) Access pattern disclosure on searchable encryption: ramification, attack and mitigation. In: 19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5–8, 2012. The Internet Society
- Jain AK, Ross A (n.d.) Introduction to biometrics. In: Jain AK, Flynn, Ross A (eds) Handbook of biometrics. Springer, pp 1–22. ISBN:978-0-387-71040-2
- Klinger E, Starkweather D (n.d.). phash.org: Home of pHash, the open source perceptual hash library (2008–2010), <http://www.phash.org/>
- Kreuter B, Shelat A, Mood B, Butler K (2013) PCF: a portable circuit format for scalable two-party secure computation. In: Proceedings of the 22th USENIX security symposium, Washington, DC, USA, August 14–16
- Lin D, Rao P, Bertino E, Li N, Lobo J (2010) EXAM: a comprehensive environment for the analysis of access control policies. Int J Inf Secur (IJIS) 9(4):253–273
- Lindell Y, Pinkas B (2000) Privacy preserving data mining. In: Advances in cryptology. Springer
- Liu D, Bertino E, Yi X (2014) Privacy of outsourced K-means clustering. In: Proceedings of the 9th ACM symposium on information, computer and communication security (ASIACCS), Kyoto (Japan), June 4–6
- Nabeel M, Bertino E (2014a) Privacy preserving delegated access control in public clouds. IEEE Trans Knowl Data Eng 26(9):2268–2280
- Nabeel M, Bertino E (2014b) Attribute based group key management. Transact Data Priv 7(3):309–336

- Nabeel M, Shang N, Bertino E (2013) Privacy-preserving policy-based content sharing in public clouds. *IEEE Trans Knowl Data Eng* 5(11):2602–2614
- Naor M, Pinkas B (2001) Efficient oblivious transfer protocols. In: *Proceedings of the twelfth annual symposium on discrete algorithms*, January 7–9, Washington, DC
- Naor M, Pinkas B (2006) Oblivious polynomial evaluation. *SIAM J Comput* 35(5):1254–1281
- Ni Q, Lobo J, Calo SB, Rohatgi P, Bertino E (2009) Automating role-based provisioning by learning from examples. In: *Proceedings of SACMAT 2009*, 14th ACM symposium on access control models and technologies, Stresa, Italy, June 3–5, 2009, *Proceedings. ACM*
- Ni Q, Bertino E, Lobo J, Brodie C, Karat CM, Karat J, Trombetta A (2010) Privacy-aware role-based access control. *ACM Trans Inf Syst Secur* 13(3)
- Paulet R, Kaosar MG, Yi X, Bertino E (2014) Privacy-preserving and content-protecting location based queries. *IEEE Trans Knowl Data Eng* 26(5):1200–1210
- Pedersen TP (1992) Non-interactive and information-theoretic secure verifiable secret sharing. In: *Proceedings of CRYPTO 1991*, vol 576. LNCS
- PrivBioMTAuth (2018) Privacy preserving biometrics-based and user centric protocol for user authentication from mobile phones. *IEEE Trans Inf Forensic Secur* 13(4):1042–1057
- Rao F-Y, Cao J, Bertino E, Kantarcioglu M (2019) A hybrid private record linkage scheme: separating differentially private synopses from matching records. *ACM Trans Priv Secur* 22(3):15:1–15:36
- Scannapieco M, Figotin I, Bertino E, Elmagarmid A Privacy preserving schema and data matching. In: *Proceedings of 2007 ACM SIGMOD international conference on management of data*
- Seo S-H, Nabeel M, Ding X, Bertino E (2014) An efficient certificateless encryption for secure data sharing in public clouds. *IEEE Trans Knowl Data Eng* 26(9):2107–2119
- Shamir A (1979) How to share a secret. *Commun ACM* 22(11):612–613
- Shang N, Nabeel M, Paci F, Bertino E (n.d.) A privacy-preserving approach to policy-based content dissemination. In: *ICDE 2010*
- Shebaro B, Oluwatimi O, Midi D, Bertino E (2014) IdentiDroid: android can finally wear its anonymous suit. *Transact Data Priv* 7(1):27–50
- Ulusoy H et al (n.d.) Vigiles: fine-grained access control for mapreduce systems. In: *2014 IEEE International Congress on Big Data (BigData Congress)*, pp 40–47
- Vaidya J, Zhu Y, Clifton C (2006) Privacy preserving data mining. *Adv Inf Secur* 19(Springer):1–121
- Wang HX, Nayak K, Liu C, Shi E, Stefanov E, Huang Y (2014) Oblivious data structures. *IACR Cryptol ePrint Arch* 185
- XACML. https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml.
- Yao AC (1986) How to generate and exchange secrets. In: *Proceedings of the 27th IEEE symposium on Foundations of Computer Science*, pp 162–167
- Yi X, Rao F, Bertino E, Bouguettaya A (2015) Privacy-preserving association rule mining in cloud computing. In: *Proceedings of the 10th ACM Symposium on Information, Computer and Communication Security (ASIACCS)*, Singapore, April 14–16. (in print)



Privacy Preservation in Big Data Analytics

Yu-Chuan Tsai¹, Shyue-Liang Wang² and
Tzung-Pei Hong³

¹Library and Information Center, National
University of Kaohsiung, Kaohsiung, Taiwan

²Department of Information Management,
National University of Kaohsiung, Kaohsiung,
Taiwan

³Department of Computer Science and
Information Engineering, National University of
Kaohsiung, Kaohsiung, Taiwan

Article Outline

Definition of Subject

Glossary

Introduction

Privacy Preservation Data Publishing

The Challenges on Big Data Privacy Preservation

Summary

Bibliography

Definition of Subject

This work presents a quick review of current privacy preservation techniques and points to challenges faced in the big data era. The need for privacy preservation in an information society is ubiquitous. There are well-known examples of privacy breaches and security fraud that have forced governments around the globe to pass and enforce privacy acts. More advanced privacy preservation techniques have been developed and deployed in modern information systems and apps. However, due to the volume and variety of big data arriving at an accelerated speed, current privacy preservation techniques are facing new challenges and require an uplift to deal with the urgent need for big data privacy and security.

We summarize some directions of research combining big data technology and privacy preservation techniques.

Glossary

Privacy Preservation; Data Publishing; Network Publishing; Data Mining; Re-identification; Anonymization; k-Anonymity; l-Diversity; t-Closeness; Differential Privacy

Introduction

Big data and its analysis have been one of the revolutionary research topics in recent years. According to the Gartner's 2014 Hype Cycle report, the term "big data" passed the *Peak of Inflated Expectations* state and will soon enter the *Trough of Disillusionment* state, so it continues to garner a lot of investments (Gartner 2014). Big data plays a great role in various areas, such as academic research, government, industry, healthcare, finance, business, and eventually, the society as a whole (Sagiroglu and Sinanc 2013). These data are generated from structured data, unstructured data, and semi-structured data. Structured data includes online transactions, health records, and search query logs, among others. Unstructured data consists of emails, videos, audios, images, mobile data, sensor data, posts, and click streams. Semi-structured data covers social web sites interactions and scientific data.

Depending on the estimation from IBM, every day 2.5 quintillion bytes of data are created, and 90% of this amount of data has been created in the last 2 years globally (IBM). An International Data Corporation (IDC) report predicts that "From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes (or 40 trillion gigabytes). From now

until 2020, the digital universe will about double every two years” (Gantz and Remsel 2012). Therefore, big data are difficult to capture, store, manage, share, visualize, and analyze using traditional tools.

In general, the characteristics of big data lie in three aspects: (1) data are huge and various, (2) data are not suitable for relational databases, and (3) data are generated, captured, and analyzed rapidly. One of challenges for big data analytics is that the data growth rate exceeds the ability to design a suitable platform for data analysis. Therefore, cloud computing techniques have been proposed as a solution for enterprises or organizations to construct a powerful architecture and perform large-scale, complex computations.

In January of 2010, the CEO of Facebook, Mark Zuckerberg, declared that the age of privacy is over. The rise of social network websites means that the privacy is easily exploited (Munir et al. 2015). The privacy of individuals and organizations is a big issue that needs to get a great deal of attention. Even after removing identifiers, such as names and social security numbers, the collected big data may re-identify individuals by linking data with other data or knowledge. Big data collects a massive amount of social network behavior, transaction logs, financial data, and other government sectors’ open data, and will certainly cause a loss of privacy for individuals. Therefore, big data can easily connect to individuals by combining the various sources information. Privacy concerns related to data processing, sharing, and analyzing of sensitive personal information is increasing for big data (Zhang et al. 2014a). In this era of big data, privacy preservation issues thus are gaining a great deal of attention from researchers, and it is important that the privacy challenges related to identification, predictive analysis, and data collection be taken into consideration when data is being combined from multiple of sources (Bell et al. 2014).

What Is Big Data

Big data concerns large-volume, complex, growing datasets with multiple, automatic sources. Big data can be formally characterized by the HACE theorem (Wu et al. 2013). The HACE theorem

defines that big data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data. Various areas, such as academic research, government, industry, health care, finance, business, and society as a whole, include concerns about big data. Depending on the characteristics of the big data, it includes massive amounts of data collected from structured data, such as transaction records, unstructured or heterogeneous data, and some semi-structured data. Big data is a complex process with regard to how it is perceived, acquired, managed, and shared using traditional IT techniques, and it is hard to analyze with real-time analytics (Chen et al. 2014).

According to the previously mentioned characteristics, it is easy to describe the characteristics of big data. The first definition of big data was proposed by Laney, who defined a $3V$ model in 2001 (Laney 2010). Recently, big data is regarded as various views of $3Vs$ or $4Vs$ (Chen and Zhang 2014; Chen et al. 2014; Hashem et al. 2015; Savitz 2012; Zikopoulos et al. 2011). Gartner, IBM, and Microsoft and other companies are using the $3Vs$ model to describe big data (Chen et al. 2014). The most well-known definition of big data was proposed by Gartner in 2012, “Big data are high-volume, high-velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization,” which was characterized by three Vs: *volume*, *velocity*, and *variety* (Savitz 2012). *Volume* refers to the sizes of datasets, such as the amount of all types of data sources, which are huge. *Velocity* refers to the speed of data transfer, such as input and output. *Variety* refers to heterogeneous and diverse dimensional data, where the data types and sources are varied, such as the different types of data collected from sensors, IoT devices, social networks, data logs, text, video, and so on, in structured, semi-structured, or unstructured formats. The fourth V can be *value*, *variability*, or *virtual*, and is usually extended to include *veracity*, which refers to the biases, noises, and abnormality in data (Chen and Zhang 2014; Chen et al.

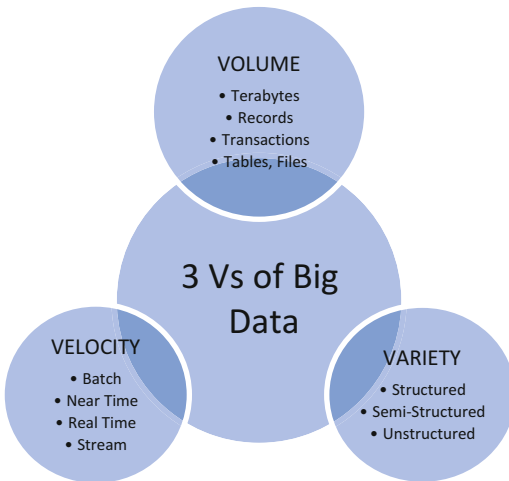
2014; Hashem et al. 2015). Figures 1 and 2 show the properties of 3Vs and 4Vs for big data, respectively.

Open Data for Governments

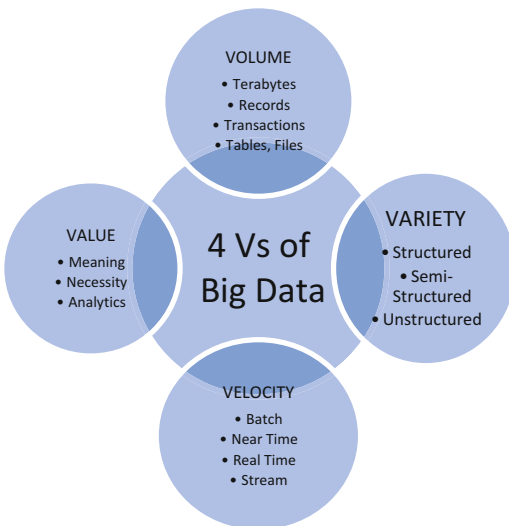
Many governments such as the United States, Australia, the European Union, the UN, and India pay a great deal of attention to big data and

analytics. In March of 2012, the Obama administration announced a \$200 million dollars investment to the project “Big Data Research and Development Initiative” (Weiss and Zgorski 2012). It was launched in six Federal departments and agencies, which have invested more than \$200 million in research and development investments to the National Science Foundation (NSF), the National Institutes of Health (NIH), the Department of Defense (DoD), the Department of Energy, and the United States Geological Survey. The NSF and the NIH will support the development of the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from massive datasets. The NIH announced that the world’s biggest dataset on human genetic variation is freely available on Amazon’s computing cloud-AWS. The DoD created a project “placing a big bet on big data” and invests 250 million dollars annually across the department of the military in a series of programs. The Department of Energy and the US Geological Survey also announced involvement in big data analysis, management, and visualization projects. The Open, Public, Electronic and Necessary (OPEN) Government Data Act provides a government-wide mandate for federal agencies to publish their information as open data and uses the standardized, machine-readable data formats. The OPEN Government Data Act was passed at the US House of Representatives on November 2017. In January 2019, the OPEN Government Data Act became a law and made Data.gov a requirement in statue (Open, Public, Electronic, and Necessary Government Data Act 2019). Data.gov easily provides access to government datasets, which covers a huge number of topics, such as weather, demographics, health, education, climate, consumer, energy, and finance.

The Australian government published the report “Australian Public Service Better Practice Guide for Big Data” in January of 2015 (Australian Government 2015). In this report, the Australian government and agencies integrate the big data solutions into research, delivery, management, and reporting. It also discusses the privacy impact for open data on the public sectors. It



Privacy Preservation in Big Data Analytics,
Fig. 1 Three versus of Big Data



Privacy Preservation in Big Data Analytics,
Fig. 2 Four versus of Big Data

expresses interest in improving big data competence, which includes identifying business requirements for big data, developing capabilities for cloud computing as well as the necessary infrastructure to accompany it, identifying high value datasets, advising the government's use of third-party datasets, and promoting privacy designs and privacy impact assessments.

In 2010, the EU started planning the construction of the infrastructure of the eGovernment. The EU governments are devoted to constructing open data from the open government. The reports on open data for the EU show how businesses and citizens still face several difficulties in finding and re-using public sector information. The EU constructed the EU Open Data Portal, which is the single point of access to a growing range of data produced by the institutions and other bodies of the EU. Data are free to use, reuse, link, and redistribute for commercial or noncommercial purposes. In 2013, the United Nations issued the Big Data for Development report, which explores how governments utilize big data to better serve the public, such as social media and other source information, and how it protects their populations (Chen et al. 2014; United Nations Global Pulse 2012).

Big Data: Issues and Challenges

Big data is gathered from various sources, such as internal transaction logs, the social relationships from emails, observations from internal business analysis rules, external relationships from social websites, government open data portals, mobile location information, and so on. Why does big data attract such great interest from researchers, companies, and governments? Previously, people made decisions using experience and had few data on which to base decisions. Now, people can make decisions using the results of data analyses or suggestions from data mining results. We can easily construct prediction or recommendation models to help with decision making. In the future, big data analytics can be greatly useful to make decisions. According to the results of big data analytics, it is possible to get closer to the truth and

show hidden information in these data. Analytics can also help to create useful models for business, government, or help to provide recommendations related to other issues. In many areas, such as academic research, government, industry, health care, finance, business, entertainment, science, and the society at large, big data play an evaluative role (Sagiroglu and Sinanc 2013). According to the characteristics of big data and its applications, big data is a revolution on data analysis, prediction, and prescription. There are some issues and challenges related to big data from different viewpoints, however.

- (A) Data storage – scalability, integration, and infrastructure: The big data storage framework must be flexible enough to store large amounts and different types of data. It is difficult to achieve this using the traditional IT framework in terms of storage scalability, data cleaning, and data integration. Big data builds on distribution and a heterogeneous environment. Therefore, it is difficult but necessary to determine how to adapt to the rapid and continuous increases in data, how to integrate data in the distribution environment, and how to construct a cost-effective and efficient storage framework.
- (B) Data process – analysis, data quality, and management: The big data analysis has big challenges related to data cleaning and integration before analysis. The computational cost of big data analysis is important to take into consideration in a distributed analysis environment. It is difficult to determine how to merge these massive and varied data sources. The integration problem has to be dealt with in situations, such as data inconsistent, missing data, and bad links, before doing big data analysis. Determining how to respond analytical results in real or near time is important for big data analytics. Dimensional reduction methods and sampling methods are typical methods used to reduce the input data volume and to accelerate the big data analytical process.

- (C) Government: The benefits of big data for society and governments will be great, which can be part of solutions used to solve global problems that include things like crime prevention, climate change, economic models development, and so on (Munir et al. 2015). Determining how to fragment the original data collected from the government sectors is becoming more and more critical for government open data. It is important to arrive at methods by which to deliver open data, which can be reused for other objectives. However, it is also time to review the laws and make them suitable for the era of big data.
- (D) Privacy risk: Big data carries significant privacy, security, and transferring risks which will continue to escalate (Bell et al. 2014). There has to be a balance between the opportunities and challenges and the benefits and the risks of identity disclosure (Bell et al. 2014; Munir et al. 2015). In big data analytics, linking attacks related to privacy breaches are widespread. Privacy breaches on big data analytics, such as automated recommendation systems, predictive analysis for business models, crime prevention and disease epidemics have to be addressed. In big data analytics, many different types and sources of datum need to be anonymized. However, anonymity is insufficient for privacy preservation and is necessary and difficult to achieve in practice. The problems of re-identification attacks are complex issues that are hard to prevent.

Most of the works on this topic have been devoted into the benefits of big data analytics. Although privacy breaches is an open issue, big data analytics is being faced with find methods to protect privacy and data and enforce the privacy protection laws. In (Xu et al. 2014), big data was divided into four major scenarios, such as data provider, data collector, data miner, and decision maker, in order to examine privacy preservation in big data analytics. The data provider is concerned about whether he/she is providing sensitive data that should not be accessed by a data collector.

Data collectors are concerned about releasing collected data and with providing high privacy and high utility when releasing data, so they are interested in the research topic, *privacy preserving data publishing* (PPDP). The data miner is concerned about the useful results of data mining, which makes use of released data, so it focuses on *privacy preserving data mining* (PPDM). The decision maker gets the results from data miners and transmits them into useful information to make decisions. In this work, we focus on the PPDP and PPDM research topics for big data analytics.

Privacy Law

Through the years, in order to protect personal data from public breaches, national laws on freedom of information and on-line privacy protection acts have been proposed and implemented in many countries. In (Banisar 2006), the European Union Parliament passed Directive 94/46/EC, which forbids sharing data with states that do not protect privacy, which was effective starting in October of 1998. The United States passed the HIPAA for healthcare data in 1996, the Crann-Leach-Bliley Act in 1999 for financial institutions, and the COPPA for children's online privacy in 2000. Canada passed the PIPEDA 2000 for personal information protection and electronic document act in 2004. The United States federal law called The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) was enacted in 2002 and was intended to determine which companies and statistical agencies should not be exposed in ways that lead to inappropriate or surprising identification of individuals. However, even though such laws can deter or punish offenders, they cannot completely stop privacy infractions from publicly available information. Technologies such as privacy-preserving data publishing, privacy-preserving network publishing, and privacy-preserving data mining that can help to preserve the privacy of personal or institutional information before data are published to the public should be developed to complement the laws on freedom of information.

Big data has great challenges related to privacy protection. Privacy risks are generated from

automatic data linkages among various seemingly non-identifiable data sources. However, these data linkages result in privacy breaches for individuals. In 2014, the White House reviewed big data policies, including privacy issues (Executive Office of the President 2014). In this report, it was concluded that big data can be used to cause social harm and that efforts must be made ensuring an appropriate balance between individuals and institutions. Some countries do not have national privacy laws or laws specific to big data. However, they have extended existing laws, such as some national laws on freedom of information and online privacy protection, to make them suitable for the privacy issues related to big data (Bell et al. 2014; Executive Office of the President 2014). Therefore, these laws can respond to big data collection. The Federal Trade Commission (FTC) revised the COPPA requirements in 2013 to include that third parties can recognize a user across different sites or online services (Bell et al. 2014; Executive Office of the President 2014). In 2012, the Obama Administration released the Consumer Privacy Bill of Rights as other countries and international organizations began to review their own privacy frameworks (Executive Office of the President 2014; Monreale et al. 2014). The European Union (EU) privacy laws apply to information, which alone, or together with other information acquired by a data controller, can identify living individuals. In 2012, the European Commission proposed a major revision of the existing legislative framework on the protection of personal data, which is established by Directive 94/46/EC (European Commission 2014). This proposal enhances two distinct rights, the right to data portability and the right to be forgotten. On May 13, 2014, the Court of Justice of the European Union applied this rule, the right to be forgotten, to requiring Google to remove a specific personal search results data about a Spanish person in 1998. In 2014, the EU Data Protection Directive 95/46/EC adopted a national provision to protect individuals in regard to processing and free movement of

personal data. In May 2018, the EU applied a new data protection regulation, the General Data Protection Regulation (*GDPR*), which replaced the Data Protection Directive 95/46/EC as the primary law that regulates how companies protect all EU citizens' personal data (European Commission 2018). *GDPR* is based on eight principles, which includes collection limitation principle, data quality principle, purpose specification principle, use limitation principle, security safeguards principle, individual participation principle and accounting principle, and strengthen the use of five rights, such as right of access, right to restriction of processing, data portability, right to object, and right to be forgotten. In the *GDPR*, the key privacy and data protection requirements are as follows: (1) requiring the agreement for personal data collection and processing; (2) anonymizing collected data to protect personal privacy; (3) providing personal privacy breach notifications. Privacy-preserving data publishing and privacy-preserving data mining technologies can also help to protect personal privacy, when the public or mined data results are published or applied to a business benefit model. In June 2018, California passed a consumer privacy act, AB 375 (California Consumer Privacy Act 2018). California Consumer Privacy Act (CCPA) goes on effect on January 1, 2020. It protects California residents regarding collection, use, disclosure, and sale of their personal information. CCPA applies to larger companies and data brokers and is the most rigorous privacy law in the United States, with more impact on US companies than *GDPR*.

Privacy Preservation Data Publishing

In the big data collection scenario, data providers may provide sensitive information with individuals in released data (Xu et al. 2014). Before data publishing, the original data has to be anonymized, and the sensitive information has to be removed from the released data to prevent

privacy breaches. Privacy preservation in regard to published data, such as market-based data, recommendation data, search query log data, and social network data, has attracted a great deal of interest due to breaches of privacy that can occur when data are published. The published data, which are released for research or to satisfy government regulations, concern issues of privacy, such as re-identification attacks, that are associated with personal identities, medical diseases/surgeries, political/religious/sexual preferences, transactions related to purchased items, or personal information found on social website profiles. Privacy is a great concern when information is published and shared. Privacy and utility are the trade-off issues in the topic of PPDP. According to the big data collection scenario, published data can be divided into three types of data: relational, transactional, and social network data.

Well-Known Examples for Privacy Preservation in Re-identification Problems

Current practices for protecting user privacy in published data rely on removing all identifiable personal characteristics such as actual names, social security numbers, or IP addresses, limiting access, “fuzzing” the data, eliminating unnecessary groupings, and augmenting with additional data. However, the publishing of such data makes the target still vulnerable to an attack that identifies the specific individual through performing different kinds of structural, nonstructural, and linking attacks.

The well-known example of privacy preservation is that an attack can re-identify a specific individual from relational published data. The given public voter registration data and private microdata that identifying information (name and social security number) had been removed in the patient data of Massachusetts’s state employees. This was a simple “linking” attack by joining two datasets, private patient data and the public voter registration data to re-identify the individual and medical history of the state’s governor. In fact, according to one study, approximately 87% of the population of the United States can be uniquely identified on the basis of their 5-digit zip code, sex, and date of birth

(Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002).

For published transactional data, *America Online (AOL)* released a large portion of its search engine query logs for research purposes in August of 2006. The dataset contained 20 million queries posed by 650,000 *AOL* users over a 3 months period. Before releasing the data, *AOL* replaced each user’s name with a random identifier. However, by examining unique query terms without any sensitive information, the *New York Times* (Barbaro and Zeller Jr 2006) demonstrated that searcher No. 4,417,749 could be traced back to Thelma Arnold, a 62-year-old widow who lived in Lilburn, Georgia, at the time.

Recently, there have many online social network websites established, which have already attracted many users to create and publish their profiles with content and also to create links to other users with whom they associate. These websites have their own privacy policies and allow members to control the disclosure of their personal information. However, this is still not enough to protect individual privacy. For example, John is a member of an online social website, and currently lives in New York City. He sets his privacy policy to not publicly release his current city. However, some of John’s friends, who also live in New York City, do not set their privacy policies with their current city hidden. An adversary can use linked-structure information to derive which city John lives in, and then John’s privacy is breached.

Techniques for Privacy Preserving Data Publishing

Many privacy models and anonymization approaches have been studied (Fung et al. 2010). Privacy preserving data publishing is intended to prevent re-identification attacks. According to motivated examples, published data had been de-identified by hiding the identifying information of the individuals, such as names and Social Security Numbers (SSN), so that the identities of individuals to whom the data referred could not be disclosed, as shown in Fig. 3.

The dataset has been divided into three types of attributes, such as identifier attributes, *Quasi-*

User ID	Name	DOB	Sex	Zip code	Disease
111	Andre	1/21/76	Male	53915	Heart Disease
222	Beth	4/13/86	Male	53605	Hepatitis
333	Corel	4/28/84	Male	53503	Bronchitis
444	Dan	1/21/76	Female	53903	Broken Arm
555	Ellen	4/23/86	Male	53706	Flu
666	Sara	1/28/76	Female	53936	Hang Nail
Identifiers (Hidden)	Identifiers (Hidden)	<i>Quasi-Identifiers</i> attributes			Sensitive attributes (Private)

Privacy Preservation in Big Data Analytics, Fig. 3 Hospital Patient data (ID and name are hidden)

Identifiers attributes, and sensitive attributes, as shown in Fig. 3.

- (a) Identifier attributes: Attributes that can identify an individual person, for example, User ID, Name, or SSN. This type of attribute has to be hidden in published data.
- (b) *Quasi-Identifiers (QI)* attributes: A collection of attributes, with the ability to link to other information, and can be used to identify an individual, for example, zip code, age, and date of birth. The published data needs to anonymize this type of attribute, for which there are some techniques that can be used to do this, for example, generalization, suppression, perturbation, and permutation.
- (c) Sensitive attributes (*SA*): Attributes, which users keep as private, do not be disclosed, for example, disease.

In order to prevent re-identification attacks based on the published data, the anonymized techniques are most widely used. The generalization, suppression, perturbation, and permutation methods have often been used to achieve anonymization, for example, *k-anonymity* (Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002), *l-diversity* (Islam and Brankovic 2011; Machanavajjhala et al. 2007), and *t-closeness* (Li et al. 2007). Figure 4 shows the 3-anonymity, which uses the generalization and

suppression methods, for the original dataset in Fig. 3.

In this work, we discuss three different types of data with data publishing, such as relational, transactional, and social network data to be anonymized. To address the prevention of the re-identification problem, the most widely studied anonymity paradigm is *k-anonymity* (Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002) and its variants, *l-diversity* (Islam and Brankovic 2011; Machanavajjhala et al. 2007), and *t-closeness* (Li et al. 2007). Differential privacy (Dwork 2006; Dwork et al. 2006) provides formal privacy intended to guarantee the outcome of computation, which is insensitive to any particular individual. In Table 1, we show a summary of the anonymization methods on different data types of data for privacy preservation in data publishing.

k-Anonymity on Relational Data

In previous works (Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002), the anonymity model, *k-anonymity* model, has been proposed, in which each record in a published dataset is required to be indistinguishable among at least *k-1* other records with respect to the set of *QI* attributes, to prevent the problem of re-identification attacks. Depending on such an equivalence class or anonymized group, there have been many techniques proposed:

DOB	Sex	Zip code	Disease
1/**/76	Human	[53900, 53940]	Heart Disease
4/**/8*	Male	[53500, 53800]	Hepatitis
4/**/8*	Male	[53500, 53800]	Bronchitis
1/**/76	Human	[53900, 53940]	Broken Arm
4/**/8*	Male	[53500, 53800]	Flu
1/**/76	Human	[53900, 53940]	Hang Nail
<i>QI</i> attribute (suppression)	<i>QI</i> attribute (generalization)	<i>QI</i> attribute (generalization)	Sensitive attributes (Private)

Privacy Preservation in Big Data Analytics, Fig. 4 3-anonymized for original Hospital Patient dataset

generalization, suppression, clustering, permutation, perturbation, etc. This generalization process of data anonymization is called *recoding*. A particular value is always mapped to the same generalized value in the entire dataset, which is called *global recoding*. However, a particular value in different anonymized groups is mapped to different generalized values in the entire dataset, which is called *local recoding* (Aggarwal et al. 2006). To achieve *k-anonymity* through generalization, most of these works concentrated on an additional assumption called *Domain Generalization Hierarchies (DGH)* (Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002) on each *QI* attribute or did not assume the *DGH* (Aggarwal et al. 2006; Meyerson and Williams 2004; Wang et al. 2011b).

The suppression anonymization process attempts to remove certain attribute values (cell suppression) (Aggarwal et al. 2005) or entire records (tuple suppression) (Samarati 2001) from the dataset. The suppression method can be combined with the generalization technique (Machanavajjhala et al. 2007) or used independently in approximation algorithms (Park and Shim 2007). Regardless of whether generalization (Liu and Yang 2011) or suppression (McAfee and Brynjolfsson 2012; Ni and Chong 2012) approaches are being used, it has been shown that *k-anonymization* on relational data is

NP-hard. Approximation algorithms with ratios of $O(k \cdot \log k)$ (McAfee and Brynjolfsson 2012), $O(k)$ (Aggarwal et al. 2005), and $O(\log k)$ (Ni and Chong 2012) that minimize the number of entries suppressed on relational data were subsequently proposed. However, due to the loose approximation ratio, the algorithms show poor performance for large *k*.

The *permutation*-based anonymization process deals with the *SA* from their associated *QI* attributes (Xiao and Tao 2006). In (Xiao and Tao 2006), the *Anatomy* algorithm published *QID* attributes directly but permuted the *SA* values among the records. In (Agrawal and Srikant 2000), random *perturbation* by adding noise to the data was proposed to prevent re-identification of records. However, adding random noises is correlated with the original data (Huang et al. 2005), and therefore outliers can be easily exposed when an attacker has access to external knowledge (LeFevre et al. 2006). The proposed method, *DETECTIVE*, which is combined with clustering categorical values and then used to add noise into all attributes, maintains high data utility (Islam and Brankovic 2011). According to the distance-preserving and distribution-preserving perturbations, it maintains the characteristic of nearest neighborhood (Ni and Chong 2012).

Privacy Preservation in Big Data Analytics, Table 1 Summary the anonymization methods on different data types

Data Source	Privacy Content	Anonymization Methods	Description
Relational	Identity Quasi-Identifiers (<i>QI</i>) attributes Sensitive attributes (<i>SA</i>)	<i>k</i> -anonymity	Generalization with <i>DGH</i> on each <i>QI</i> attribute (Samarati 2001; Samarati and Sweeney 1998; Sweeney 2002)
		<i>k</i> -anonymity	Generalization without <i>DGH</i> on each <i>QI</i> attribute (Aggarwal et al. 2006; Wang et al. 2011b)
		<i>k</i> -anonymity	Cell suppression in approximation algorithm (Aggarwal et al. 2005; Park and Shim 2007)
		<i>k</i> -anonymity	Tuple suppression with <i>DGH</i> (Samarati 2001)
		<i>k</i> -anonymity	Suppression in approximation algorithm with approximation ratio $O(k \cdot \log k)$ (Meyerson and Williams 2004)
		<i>k</i> -anonymity	Suppression without <i>DGH</i> into <i>l</i> -diversity (Machanavajjhala et al. 2007)
		<i>k</i> -anonymity	Clustering-based local recoding, <i>r</i> -gathering, <i>r</i> -cellular, cluster centers are only published (Aggarwal et al. 2006)
		<i>k</i> -anonymity, <i>l</i> -diversity	<i>Anatomy</i> , permuted the <i>SA</i> values, but the <i>QI</i> attributes are published directly (Xiao and Tao 2006)
		<i>k</i> -anonymity	Random perturbation by adding noise (Agrawal and Srikant 2000; Huang et al. 2005; LeFevre et al. 2006)
		<i>k</i> -anonymity	<i>DETECTIVE</i> , combined with clustering categorical values and adding noise (Islam and Brankovic 2011)
		<i>k</i> -anonymity	Using distance-preserving and distribution-preserving perturbations (Ni and Chong 2012)
		<i>k</i> -anonymity, <i>l</i> -diversity	Suppression without <i>DGH</i> into <i>l</i> -diversity (Machanavajjhala et al. 2007)
		<i>k</i> -anonymity, <i>t</i> -closeness	Generalization with <i>DGH</i> into <i>t</i> -closeness (Li et al. 2007)
		<i>k</i> -anonymity, differential privacy	ϵ -differential privacy, guarantee the privacy concern (Dwork 2006; Dwork et al. 2006)
Transactional	Identity Quasi-Identifiers items (<i>QID</i>) Sensitive items (<i>SI</i>)	<i>k</i> -anonymity	Correlated <i>QI</i> items are permuted to a group, corresponding <i>SI</i> are made to be homogeneous (Ghinita et al. 2008; Ghinita et al. 2011)
		<i>k</i> -anonymity	Using bitmap to group similar <i>QI</i> items (Ghinita et al. 2011; Wang et al. 2011b)
		<i>k</i> -anonymity	<i>k</i> -anonymity in Approximation algorithm with approximation ratio $O(\log k)$ without <i>DGH</i> (Wang et al. 2010, 2011a)
		<i>k</i> -anonymity	Two-phase suppression anonymization in Approximation algorithm with approximation ratio $O(k \log k)$ without <i>DGH</i> (Matwin 2013)
		k^m -anonymity	Top-down local recoding generalization with <i>DGH</i> (He and Naughton 2009)
		k^m -anonymity	Top-down global recoding generalization with <i>DGH</i> and suppression (Liu and Wang 2010)
		k^m -anonymity	Bottom-up local recoding generalization with <i>DGH</i> (Terrovitis et al. 2008, 2011)
Social Network	Node identity	<i>k</i> -anonymity	Grouping nodes and labels into the equivalence class on bipartite graph (Bhagat et al. 2009)
	Label identity	<i>k</i> -anonymity	Generalization by clustering and neighborhood characteristic (Hay et al. 2008)
	Structure identity	<i>k</i> -degree	Given any node has at least <i>k</i> -1 other nodes with the same degree (Liu and Terzi 2008)

(continued)

Privacy Preservation in Big Data Analytics, Table 1 (continued)

Data Source	Privacy Content	Anonymization Methods	Description
	Node identity Label identity Structure identity	<i>k-degree-label anonymous</i>	Used three different criteria, labels on the nodes, degree of nodes, and the labels on the edges, to achieve three levels of privacy protection (Yuan et al. 2010)
	Structure identity	<i>k-automorphism</i>	Each node has at least $k-1$ other nodes with the same structure (Zou et al. 2009)
	Structure identity	<i>k-isomorphism</i>	Forming k pair-wise isomorphic sub-graphs to prevent identifying specific sub-graphs (Cheng et al. 2010)
	Structure identity	<i>k-anonymity, l-diversity</i>	Generalized node labels and added edges to prevent the 1-neighborhood attack (Zhou and Pei 2008)
	Structure identity	<i>k-possible anonymity</i>	Using the generalization method to modify the edges and weights and prevent the weighted-based attack (Liu and Yang 2011)
	Shortest path identity	Edge weight anonymity	Preserving the shortest path characteristic and minimizing to modify the edge weights (Das et al. 2010; Liu et al. 2009)
	Shortest path identity	<i>k-anonymous edge weight</i>	Preserving the shortest path characteristic and greedy-based perturbation method to minimally modify edge weights (Liu et al. 2010)
	Shortest path identity	1-neighborhood- d -radius privacy model	Minimizing the computing cost, preventing the neighborhood attack and preserving shortest distances characteristic (Gao et al. 2011)

l-Diversity and *t*-Closeness on Relational Data

Privacy and utility are the trade-off in the *k*-anonymity models. When the value of *k* increases, the privacy also increases, but the utility decreases (Punagin and Arya 2017). The *k-anonymity* model has strong privacy protection, which reduces the diversity on *SA* values in the anonymized dataset. The *l*-diversity, which gives stronger privacy guarantees on the *k*-anonymity model, has *l*-diverse in each sensitive attribute equivalence class (Machanavajjhala et al. 2007). Another privacy model, *t*-closeness, prevents a skewness attack (Li et al. 2007). The distribution of a sensitive attribute in the specific equivalence class is very different from the distribution of the overall dataset, so adversaries can re-identify a specific individual using only a little background knowledge. The *t*-closeness model is defined by a distance not larger than *t* between the distribution of an *SA* in a particular equivalence class and the distribution of the overall dataset.

Anonymization Methods on Transactional Data

The *k-anonymity* on transactional data, which uses the similar definition to that of the relational data,

requires that each transaction in a dataset has at least $k-1$ other indistinguishable transactions. According to the characteristics of transactional data, it is regards to a lack of specific quasi-identifiers (*QI*) attributes. In transactional data, all items are treated as quasi-identifiers (*QID*), which are high dimensionality and sparse. Therefore, the *k-anonymity* methods for the traditional relational data are not suitable for direct using to transactional data.

The techniques used to anonymize relational data, such as *generalization*, *suppression*, *clustering*, *permutation*, and *perturbation*, which have some extended methods for transactional data. Many bottom-up (Terrovitis et al. 2008, 2011) and top-down (He and Naughton 2009) traversals on the *DGH* for finding the minimal generalized values have been proposed for transactional data anonymization. The *k^m-anonymization* model (Terrovitis et al. 2008, 2011), which was combined with the bottom-up local recoding generalization method and the *DGH* method, achieved the anonymization of transactional data. Given the *DGH* on items, a local recoding, top-down generalization approach has been proposed to achieve

k^m -anonymity (He and Naughton 2009). Given the *DGH* on items, (Liu and Wang 2010) was a top-down generalization approach for global recoding and integrated generalization and suppression to achieve k^m -anonymity for transactional data. In (Motwani and Nabar 2008), a two-phase suppression anonymization approach without using *DGH* on transactional data was proposed. This approach firstly transformed the transactional data into binary relational data and applied the suppression anonymization on the dataset. In order to obtain an anonymization result, the second phase used the flipping concept to the suppressed data.

The *permutation*-based anonymization process deals with the *SI* from their associated *QI* (Ghinita et al. 2008, 2011; Motwani and Nabar 2008). In (Ghinita et al. 2008, 2011), the *QI* items were grouped by using the band matrix and *Gray* encoding, and the *QI* and *SI* data were published separately. The correlated *QI* items were permuted to a group, and the corresponding *SI* were made homogeneous. The problems of the suppression-based anonymization on transactional data are *NP*-hard (Motwani and Nabar 2008; Wang et al. 2010, 2011a). In (Motwani and Nabar 2008; Wang et al. 2010, 2011a), they extended the *approximation algorithms* on relational data to transactional data. In (Motwani and Nabar 2008), an *approximation algorithm* and suppression strategy were applied, for which the approximation ratio was $O(k \cdot \log k)$, which was proposed to achieve better performance for k -anonymity for of transactional data. In (Wang et al. 2010, 2011a), they concentrated on directly applying an *approximation algorithm* for k -anonymity on transactional data without using the *DGH*, for which the approximation ratios were $O(\log k)$. The *permutation-based* approaches on transactional data with sensitive items (Ghinita et al. 2008, 2011) can reduce not only time complexity but also privacy.

Anonymization Methods on Social Network Data

Social network data is made up of a social structure, which includes entities that represent the individuals that are connected by one or more attributes and edges that represent the

relationships or interactions between entities (Hinds and Lee 2008). The edge weights can represent the strength of the relationships, in which positive edge weights express the trust relationship, and the negative edge weights express bad relationships. According to the characteristics of a social network, there are three types of sensitive information, including node information, link information, and edge weight information, which should be kept private and can be used to attack. Privacy is achieved by modifying the social network graphs through adding or deleting nodes/edges before publication in order to protect the node identities or link relationships among a group of nodes.

After removing the identifier for each node, an adversary also possibly gets information based on whether edges exist or not between some specific target nodes (Backstrom et al. 2007). Therefore, only removing identifying information is not enough to prevent privacy breaches. Most of the previous works on this topic have concentrated on preserving the privacy of node identities, link structures, and sub-graph patterns (Bhagat et al. 2009; Hay et al. 2008; Li and Shen 2010; Liu and Terzi 2008; Yuan et al. 2010; Zhou and Pei 2008; Zou et al. 2009). In social network data, many types of structural attacks, such as degree-attacks, sub-graph attacks, 1-neighborhood attacks, and hub-fingerprint-attacks, have been proposed (Gao et al. 2011; Liu and Terzi 2008; Zhou and Pei 2008).

To protect a node's label information, generalization and suppression-based k -anonymity techniques for relational and transactional data can be applied on it directly. In (Bhagat et al. 2009), a grouping approach was proposed by grouping nodes and labels into classes on bipartite graphs to prevent node re-identification. In (Hay et al. 2008), a clustering algorithm was proposed that used the neighborhood characteristic, which partitioned the nodes and summarized the graph at the partition level in order to prevent privacy leakage. To protect link information, some privacy models address various types of structural attacks, such as k -degree (Liu and Terzi 2008), k -automorphism (Zou et al. 2009), and k -isomorphism (Cheng et al. 2010). In (Liu and Terzi

2008), the *k-degree* anonymity model was defined, in which there are at least $k-1$ other nodes with the same degree for any given node in the graph, and it was found to prevent re-identification of individuals. In (Zhou and Pei 2008), an algorithm was proposed which generalized node labels and added edges until each node had at least $k-1$ other nodes of 1-neighbor, with a stronger constraint to prevent the 1-neighborhood attack. In (Yuan et al. 2010), three levels of privacy models were proposed, the *k-degree-label anonymous model*, to achieve personal privacy requirements, which prevented attacks using the degree of nodes and edge weights adjacent to nodes as the background knowledge and used three different criteria, such as labels on the nodes, degree of the nodes, and the labels on the edges, for measurement. In (Zou et al. 2009), a stronger privacy model called *k-automorphism* was proposed, in which each node had at least $k-1$ other nodes with the same structure on the graph. In (Cheng et al. 2010), an *NP-hard* problem was solved, which proposed a privacy model, *k-isomorphism*, which was anonymized by forming k pair-wise isomorphic sub-graphs, which was found to be both sufficient and necessary for privacy protection.

To protect sensitive edge weight information for shortest paths, three types of works have been proposed on weighted graphs (Das et al. 2010; Gao et al. 2011; Liu et al. 2009, 2010). The first type of works attempts to preserve the shortest path characteristic between pairs of sources and destination vertices, such as the original shortest path after the anonymizing process remains the shortest path with minimal modifications (Das et al. 2010; Liu et al. 2009). The second type of works attempts to preserve the privacy of edges emitted from a given vertex (Liu et al. 2010). In this work, Gaussian randomization perturbation and greedy perturbation techniques that minimally modify the edge weights without adding or deleting any vertices and edges have been proposed. The third type of works studies the shortest distance computed in the cloud which aims at preventing outsourced graphs from neighborhood attacks. The adversary in an outsourced server cannot calculate the shortest path or

shortest distance between neighboring nodes (Gao et al. 2011). The other works on weighted graphs were based on the edges generalization approach, *k-possible anonymity* (Liu and Yang 2011), to achieve generalized anonymization groups on weighted graphs.

Privacy Versus Utility

Privacy and utility are the trade-off in the privacy preservation problem. After the anonymizing process, published data will have reduced utility, such as information loss. Higher privacy thus means lower utility of anonymized data. Many researchers are devoted to achieving a balance between these two. There are many anonymization approaches that attempt to minimize changes between the original and anonymized data and use measurement metrics to evaluate the utility loss in the anonymized data (LeFevre et al. 2006; Li et al. 2007; Machanavajjhala et al. 2007; Xu et al. 2006). It is a big challenge to quantify the information loss or data utility in anonymized data. Some metrics are used to measure information loss, such as *Normalized Certainty Penalty (NCP)*, and *KL-divergence*.

Differential Privacy

In the last few years, many researchers have been devoted to using *differential privacy* to providing provable privacy guarantees. The *differential privacy* is a disclosure metric that eliminates the re-identification risk of *k-anonymity*, *l-diversity* or their extensions of anonymized datasets, in which the given query results with and without inclusion the specific individuals, so that it ensures the privacy guarantee to the given released datasets (Dwork 2006; Dwork et al. 2006). The basic concept of *differential privacy*, in which the anonymized dataset has almost no effect on any public outcome, can be regarded as a guarantee of privacy (Heffetz and Ligett 2014). *Differential privacy* is defined as a mechanism for released data M that guarantee ϵ -*differential privacy* if for every pair of two inputs, $D1$ contains the sensitive query item t , and $D2$ does not contain the sensitive query item t , the probability of M with $D1$ should be close to the probability of

M with $D2$ (Dwork 2006; Dwork et al. 2006). *Differential privacy* provides the strongest privacy protection without based on any attackers' background knowledge. Many techniques (Friedman and Schuster 2010; Hu et al. 2015; Jagannathan et al. 2012; Li et al. 2012; Mohammed et al. 2011; Yang et al. 2012) have been proposed for applying *differential privacy* to privacy preservation data publishing and data mining. In (Friedman and Schuster 2010; Hu et al. 2015; Jagannathan et al. 2012; Mohammed et al. 2011), attempts were made to adapt differential privacy to decision trees. In (Li et al. 2012), *differential privacy* was applied to frequent itemset mining. In the tutorial (Yang et al. 2012), it introduces some techniques for *differential privacy* on data publishing.

The Challenges on Big Data Privacy Preservation

In the real world of big data applications, data is related to sensitive information, such as banking transactions, healthcare information, and some datasets with privacy issues. The *KPMG* report identified five privacy challenges that must be addressed to ensure proper control of a big data program, in which re-identification risk is one of the privacy challenges (Bell et al. 2014). The report in 2010 from the Office of the Privacy Commissioner of Canada's Consultations on Online Tracking, Profiling and Targeting and Cloud Computing suggests that information collected by behavioral advertising constitutes "personal information" without an investigation. *Netflix* and *Target* are well-known examples for privacy breaches in big data analytics.

For published recommendation data, *Netflix* announced a \$1-million prize for improving their movie rating and recommendation system in October of 2006. A dataset of 100 million movie ratings posted by 500,000 subscribers over a 6-year period was released (Amatriain and Basilico 2012). *Netflix*'s published data replaced each username with a random identifier. However, it was shown that 84% of the subscribers could be uniquely identified by an attacker who knew six

out of eight movies that the subscriber had rated outside of the top 500 movies (Narayanan and Shmatikov 2008).

A serious privacy breach on personalization recommendation, in which customers were interested in a supermarket, was shown in a report of the *New York Times* magazine in February of 2012 (The New York Times 2012). Target created a research program to investigate how shopping habits changed as a woman approached her baby's due date. According to this result, pregnant women in the first 20 weeks suddenly start buying lots of scent-free soap and extra-big bags of cotton balls. However, when a commercial coupon was sent to a father's teenage daughter that advertised baby clothes and cribs, he was angry. The coupon advertisement revealed private information that she was pregnant. Therefore, private information can be easily linked to specific individuals because commercial websites often collect the users' name, email, telephone number, and so on.

The European Union funded a four years project, EEXCESS, in 2013 as an example of preserving user privacy in recommendations by making use of intensive user profiles. In EEXCESS, there are three major objectives. Firstly, it gathers rich, high-quality content, such as social media and blogs, from educational, scientific, and cultural resources. Second, it provides a high-quality personal recommendation with users' profiles or their experiences. Finally, the privacy protection plays a large role in this project. In EEXCESS, the major challenge is that this system is cross-referenced with external big data sources and analysis of this data might generate privacy breaches. Finding a balance between creating the trust and reputation for the recommender and personal privacy preservation is a trade-off.

Current Techniques for Privacy Preservation on Big Data

Privacy preservation can be divided into two main areas. One is PPDP, and the other is PPDM. Privacy issues are concerned with linked privacy breaches on cross-referenced anonymizing datasets. The privacy preservation on data publishing focuses on data anonymization and prevention of re-identification. The common

anonymization techniques are generalization, suppression, perturbation, and permutation which are used to process anonymized published data. Privacy preservation on data mining can be divided into different topics, such as privacy protection of classification algorithms, privacy protection of association rule mining, and distributed privacy preserving collaborative recommendation. These algorithms are devoted to avoiding sharing or discourse any sensitive information inside the data. After the data anonymous process, the sensitive information is protected, but the precision of the big data analytics on the anonymous datasets is also important. Utility and privacy are also a trade-off in big data analytics.

Data publishing based on data anonymization can avoid re-identification attacks. After the data anonymization process, the published data links to different sources and types of datasets, and it cannot be identified to individuals. Based on previous works, data anonymization for privacy preservation is an *NP*-hard problem (Machanavajjhala et al. 2007; Meyerson and Williams 2004; Motwani and Narasimhan 2008; Park and Shim 2007; Wang et al. 2010, 2011a). Therefore, using data anonymization approaches on big data causes scalability and efficiency challenges because the characteristics of big data, such as 3Vs, are not suitable for extending big data anonymization (Vennila and Pryadarshini 2015). In order to deal with big data anonymization, some well-known and efficient approaches, such as the divide and conquer approach, the incremental approach, dimensional reduction, sampling, and distributed computing, have been proposed. In PPDP approaches, which were introduced in section “Privacy Preservation Data Publishing”, can be extended by horizontally splitting the data into smaller parts, such as using the divide and conquer approach, the incremental approach, dimensional reduction, sampling, and distributed computing approaches, to anonymize each part of smaller subsets in isolation in order to preserve privacy on big data. Local recoding anonymization approaches can easily be extended and applied into large datasets. Some proposed data anonymization approaches considered the high dimensional data anonymization problem (Das

et al. 2010; Gao et al. 2011; Ghinita et al. 2008, 2011; He and Naughton 2009; Terrovitis et al. 2008, 2011; Vennila and Pryadarshini 2015; Wang et al. 2011b; Xu et al. 2006). Recently, research on privacy preservation issues in the MapReduce framework on the cloud has been of great concern. In these privacy preservation issues on the cloud, there still are many open questions (Chaudhuri 2012). The cloud computing platform, such as the MapReduce framework, has been widely adopted by companies and organizations to construct the big data environment, which has become more flexible, scalable, and cost-effective (Zhang et al. 2014a). The MapReduce-based anonymization framework does not affect the quality of data fragments of the anonymized publishing data and suit for the high computation process of anonymization (Zaherzadeh et al. 2015). In (Roja et al. 2015), a two-phase top-down specialization approach was used, which was based on the generalization anonymization technique, to achieve *k*-anonymity of large-scale datasets using the MapReduce framework on the cloud. In (Amalraj and Arockiam 2015), a multi-dimensional generalization anonymization schema, which is a MapReduce based algorithm, was proposed on the public cloud. It used top-down specialization and bottom-up generalization to achieve data anonymization. In (Zaherzadeh et al. 2015), the *Mondrian* multidimensional *k*-anonymity and *l*-diversity for data anonymization was extended into the MapReduce framework and used a distributed computing process to achieve performance improvements. In (Zhang et al. 2014c), the scalability problem of sub-tree or multidimensional schemes in big data was addressed by using MapReduce. However, it used global recoding anonymization. In (Zhang et al. 2014b), the local-recoding problem was investigated for big data anonymization against proximity privacy breaches in the cloud, and there was an attempt made to gain high scalability and time-efficiency. It proposed a scalable two-phase clustering approach, which was based on MapReduce, to create high scalability using a data-parallel computation in the cloud.

Data mining is a strong technology intended to help extract hidden, unknown, new, or interesting

information from huge datasets. However, privacy preservation on data mining is applied to achieve individual privacy protection with respect to the data characteristics and different applications. Determining how to prevent sensitive information from being included in the mining results is the major concern of the proposed PPDM approaches. Many data mining approaches extend to the PPDM. In (Agrawal and Srikant 2000; Aggarwal and Philip 2008; Matwin 2013), extending PPDM approaches were surveyed in detail, such as privacy preserving association rule mining, privacy preserving classification, and privacy preserving clustering. It is important to propose or extend the original PPDM approaches to the efficient and acceptable computing time of PPDM approaches in big data analytics. Extended PPDM approaches can use incremental approach, dimension reducing, divided and conquer, sampling, and distributed computing to achieve this purpose. In (Hu et al. 2015), three basic *differential privacy* architectures were proposed, including data publication architecture, separated architecture, and hybridization of data mining and database architecture, to design privacy protection for a specific data mining algorithm, which selected a decision tree as the data mining algorithm. The privacy and utility trade-off problem are also a big challenge in the PPDM. In the future, determining how to get high privacy and good utility for the results of mining and how to mine efficiently in the cloud environment should be well addressed. In Table 2, we show a summary of the PPDP and PPDM methods for privacy preservation in big data.

Opening Issues to Privacy Preservation on Big Data

Because the characteristics of big data are quite different from those in traditional data, such as the environment, devices, and infrastructure/frameworks, the traditional methods for privacy preservation may be inefficient. In each stage of the big data lifecycle, such as collection, combination, analysis and usage, serious risk of privacy breaches for individual can crop up. Obtaining a great collection, which combines various big datasets, conducting great analysis, and getting

great use of information are big challenges. Data privacy protection laws, such as US national laws and the EU Directive, have several usage guidelines including: (1) restrictions on use, (2) notification of personal data collection and processing, (3) legitimate purpose for use, (4) insurance that the collected personal data is accurate while removing identifying information when data is used for analysis, (5) and destroying of collected data when the purpose for collection is completed (Munir et al. 2015). Several open issues for privacy preservation on big data are addressed as data publishing and data mining concerns.

Privacy Preservation Big Data Publishing Issues

The identity issue is an important opening issue in privacy preservation related to big data publishing. After data anonymization for data publishing, preventing the re-identification privacy breach is still an open issue. Privacy disclosures may still be linked to external publishing data. Big data is very complex and includes different types of source data, and it is difficult to prevent re-identification when datasets are cross-references to individuals (Xu et al. 2014). Some efficient and effective data anonymization methods have been proposed in the MapReduce framework in the cloud environment. However, scalability and computational time efficiency are still open issues.

In different application domains, data anonymity also has different limitations, such as Web search mining, Internet of Things, health care, and mobility. Data-driven data approaches for different applications intended to achieve anonymity are needed that are more flexible and require less effort than those currently available.

In the PPDP on big data, it becomes more and more difficult for an adversary to model background knowledge. However, the collected data types are becoming more and more complex, so an adversary can get background knowledge from other resources with different data types and learn more information about target-sensitive information than has been the case in the past. It is difficult to understand an adversary's resources in order to model the attacks. A comprehensive analysis model is needed in order to design a privacy model that

Privacy Preservation in Big Data Analytics, Table 2 Summary the PPDP/PPDM methods on big data environment

Data Source/ Environment	PPDP/ PPDM	Privacy Content	Scalability	Method description
Relational data	PPDP	Identity <i>QI</i> attributes Sensitive attributes (<i>SA</i>)	Yes/using vertical partition attributes method	<i>Anatomy</i> , achieve <i>k-anonymity</i> and <i>l-diversity</i> , permuted the <i>SA</i> values, but the <i>QI</i> attributes published directly (Xiao and Tao 2006)
Transactional data	PPDP	Identity Quasi- Identifiers items (<i>QID</i>)	Yes/using vertical partition and band matrix method	Correlated <i>QI</i> items are permuted to a group, corresponding <i>SI</i> are made to be homogeneous (Ghinita et al. 2008, 2011)
Transactional data	PPDP	Sensitive items (<i>SI</i>)	Yes/using vertical partition and bitmap method	Using bitmap to group similar <i>QI</i> items (Ghinita et al. 2011; Wang et al. 2011b)
Transactional data	PPDP		Yes/using the Locality Sensitive Hashing method to partition	Two-phase suppression anonymization in Approximation algorithm with approximation ratio $O(k \log k)$ without <i>DGH</i> (Matwin 2013)
Transactional data	PPDP		Yes/using the divide-and- conquer, top-down partition method	Top-down local recoding generalization with <i>DGH</i> (He and Naughton 2009)
Transactional data	PPDP		Not good/using on the limited number <i>m</i> of <i>QI</i> items	Button-up local recoding generalization with <i>DGH</i> (Terrovitis et al. 2008, 2011)
Social Network data	PPDP	Edge weight anonymity	Yes/using the linear programming model	Preserving the shortest path characteristic and minimizing to modify the edge weights (Das et al. 2010)
Social Network data	PPDP	Shortest path identity	Yes/using the approximate shortest distances within a specified error bound	Minimizing the computing cost, preventing the neighborhood attack, and preserving shortest distances characteristic (Gao et al. 2011)
Cloud Environment	PPDP	Identity <i>QI</i> attributes Sensitive attributes (<i>SA</i>)	Yes/using the MapReduce	Two-phase method by MapReduce, partition into small datasets, and anonymized datasets in parallel (Roja et al. 2015)
Cloud Environment	PPDP		Yes/using the MapReduce	Combined top-down and button-up generalization anonymization (Amalraj and Arockiam 2015)
Cloud Environment	PPDP		Yes/using the MapReduce	Two-phase top-down generalization (Zhang et al. 2014c)
Cloud Environment	PPDP		Yes/using the MapReduce proximity aware agglomerative clustering	Two-phase proximity-aware clustering method by local recoding generalization with <i>DGH</i> (Zhang et al. 2014b)
Cloud Environment	PPDP		Yes/using the MapReduce	Address the scalability problem at the previous works of PPDP and extend the previous work of <i>k-anonymity</i> , <i>Mondrian</i> , and <i>l-diversity</i> to MapReduce framework (Zaherzadeh et al. 2015)
Mixed data types/Cloud Environment	PPDM	Sensitive information	Yes/using the MapReduce	PPDM on decision tree with differential privacy (Hu et al. 2015)

prevents various kinds of attacks and to describe the behavior of adversaries.

Privacy Preservation Big Data Mining Issues

Data mining is useful for finding patterns in large collection datasets, and determining how to prevent privacy breaches is a big challenge for the PPDM on big data. The computing time efficiency of PPDM methods is also a challenge on big data analytics because big data applications, such as *Netflix* or *Amazon* recommendation systems, always need real-time or near-time response. In the cloud environment, it is necessary to extend the PPDM approaches so they have efficient analysis performance, so this is still an open issue.

In the PPDM approaches, there are two scenarios to prevent privacy disclosure. First, using anonymized data applies to the data miner and discovers the useful knowledge. This scenario ensures that nonsensitive information can be mined. However, the utility of the anonymous data becomes important and is a big challenge. It is important for future works to find a suitable way to evaluate both privacy and utility. Second, modification of sensitive data directly applies to data mining approaches and prevents sensitive disclosure. In this scenario, non-disclosure information is hidden first and then the data mining approaches are applied to discover useful knowledge. This will generate negative side effects on the results of mining. These negative/side effects need to be reduced in order to increase the utility of data mining. However, it is also an important challenge to find a suitable way to quantify the privacy and utility trade-off.

In the PPDM model, modeling a privacy attack in big data analytics is also an open issue. The complex data resources and distributed environment make it difficult to model the behavior and background knowledge related to an adversary's attacks. In the distributed cloud environment, there is no guarantee that there will be no privacy breaches when non-sensitive resources are integrated together for analysis. This is also an important challenge that needs to be solved.

Summary

In this era, big data and its analysis have become a revolutionary research topic. A great deal of attention has been generated from the various areas, such as academic research, government, industry, healthcare, finance, business, and eventually, the society as a whole will be involved. Privacy concerns about data collection, processing, sharing, and analyzing of sensitive personal information is increasing in big data and is a great challenge. After removing identifiers from the collected big datasets, these datasets may still lead to re-identification of individuals through linking with other datasets or information. For example, big data analytics merges various sources and types of big datasets. Privacy breaches are generated from automatic data linkages among various seemingly non-identifiable data sources and have become a great issue in big data analysis. Technologies such as *privacy-preserving data publishing*, *privacy-preserving network publishing*, and *privacy-preserving data mining* that can help to preserve the privacy of personal or institutional information before data are published to the public should be developed. In this work, we introduce the methods intended to prevent personal privacy breaches in published datasets on relational data, transactional data and social network data. We also introduce the methods that the *differential privacy* applies to various data anonymization methods or data mining methods.

However, traditional privacy preserving techniques are not suitable for big data because of the characteristics of big data, such as the size of the datasets, the rapid generation, capture, and analysis of datasets rapidly, and the continuous growth of datasets, so many privacy preservation concerns related to big data, which are combined with the cloud computing environment and MapReduce techniques to improve efficiency and effectiveness, were proposed. Privacy preservation is still an open issue for big data and analytics, and this issue includes making it more efficient, effective, scalable, flexible, and so on, and further will require a great deal of attention from researchers, organizations, companies, and governments. Privacy and data protection in a

legal framework are also necessary to address the high level of concern in this big data era.

Bibliography

- Aggarwal CC, Philip SY (2008) A general survey of privacy-preserving data mining models and algorithms. Springer, New York
- Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A (2005) Anonymizing tables. In: Proceedings of the 10th international conference on database theory, pp 246–258
- Aggarwal G, Feder T, Kenthapadi K, Khuller S, Panigrahy R, Thomas D, Zhu A (2006) Achieving anonymity via clustering. In: Proceedings of the 25th ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems, pp 153–162
- Agrawal R, Srikant R (2000) Privacy preserving data mining. In: Proceedings of ACM SIGMOD international conference on management of data, pp 439–450
- Amalraj I, Arockiam L (2015) Scalable multidimensional anonymization algorithm over big data using map reduce on public cloud. *J Theor Appl Inf Technol* 74(2):221–231
- Amatriain X, Basilico J (2012) Netflix recommendations: beyond the five stars. The Netflix Tech Blog
- Australian Government (2015) Australian public service better practice guide for big data. Australian Government
- Backstrom L, Dwork C, Kleinberg JM (2007) Wherefore art thou r3579x? anonymized social networks, hidden patterns and structural steganography. In: Proceedings of World Wide Web, pp 181–190
- Banisar D. Freedom of information around the world 2006 – a global survey of access to government information Laws. www.privacyinternational.org/foi/survey
- Barbaro M, Zeller Jr T (2006) A face is exposed for AOL searcher no. 4417749. *New York Times*
- Bell G, Rotman D, VanDenBerg M (2014) Navigating big data's privacy and security challenges. KPMG Advisory Institute
- Bhagat S, Cormode G, Krishnamurthy B, Srivastava D (2009) Class-based graph anonymization for social network data. In: Proceedings of Very Large Data Bases, pp 766–777
- California Consumer Privacy Act, AB 375 of June 2018. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=20170180AB375
- Chaudhuri S (2012) What next? A half-dozen data management research goals for big data and the cloud. In: Proceedings of the 31st symposium on principles of database systems (PODS), pp 1–4
- Chen CLP, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf Sci* 275:314–347
- Chen M, Mao S, Liu Y (2014) Big data: a survey. *Mobile Netw Appl* 19(2):171–209
- Cheng J, Fu A, Liu J (2010) K-isomorphism: privacy preserving network publication against structural attacks. In: Proceedings of ACM SIGMOD international conference on management of data, pp 459–470
- Das S, Egecioglu O, Abbad AE (2010) Anonymizing weighted social network graphs. In: Proceedings of international conference on data engineering, pp 904–907
- Dwork C (2006) Differential privacy. In: Proceedings of the international colloquium on automata, languages and programming (ICALP), pp 1–12
- Dwork C, McSherry F, Nissim K, Smith A (2006) Calibrating noise to sensitivity in private data analysis. In: Proceedings of the 3rd theory of cryptography conference (TCC), pp 265–284
- European Commission (2014) Progress on EU data protection reform now irreversible following European Parliament vote. http://europa.eu/rapid/press-release_MEMO-14-186_en.htm
- European Commission (2018) Reform of EU data protection rules, [online]. https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. 2018
- Executive Office of the President (2014) BIG DATA: sizing opportunities, preserving values. https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf
- Friedman A, Schuster A (2010) Data mining with differential privacy. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 493–502
- Fung BCM, Wang K, Chen R, Yu PS (2010) Privacy-preserving data publishing: a survey on recent developments. *ACM Comput Surv* 42(4):Article 14
- Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows and biggest growth in the far east. In: Proceedings of IDC iView, IDC analyze the future
- Gao J, Xu Y, Jin RM, Zhou JS, Wang TJ, Yang DQ (2011) Neighborhood-privacy protected shortest distance computing in cloud. In: Proceedings of ACM SIGMOD international conference on management of data, pp 409–420
- Gartner (2014) Gartner's 2014 Hype cycle for emerging technologies maps in journey to digital business
- Ghinita G, Tao Y, Kalnis P (2008) On the anonymization of sparse high-dimensional data. In: Proceedings of IEEE 24th international conference on data engineering (ICDE), pp 715–724
- Ghinita G, Kalnis P, Tao Y (2011) Anonymous publication of sensitive transactional data. *IEEE Trans Knowl Data Eng* 33(2):161–174
- Gopalkrishnan V, Steier D, Lewis H, Guszczka J (2012) Big data, big business: bridging the gap. In: Proceedings of the 1st international workshop on big data, streams and heterogeneous source mining: algorithms, systems, programming models and applications, pp 7–11
- Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU (2015) The rise of “big data” on cloud

- computing: review and open research issues. *Inf Syst* 47:98–115
- Hay M, Miklau G, Jensen D, Towsley DF, Weis P (2008) Resisting structural re-identification in anonymized social networks. *Proc Int Conf Very Large Data Bases* 1(1):102–114
- He Y, Naughton JF (2009) Anonymization of set-valued data via top-down, local generalization. In: *Proceedings of the international conference on Very Large Databases (VLDB)*, pp 934–945
- Heffetz O, Ligett K (2014) Privacy and data-based research. *J Econ Perspect* 28(2):75–98
- Hinds H, Lee RM (2008) Social network structure as critical success condition for virtual communities. In: *Proceedings of the 41st Hawaii international conference on systems science*, p 323
- Hu X, Yuan M, Yao J, Deng Y, Chen L, Yang Q, Guan H, Zeng J (2015) Differential privacy in telco big data platform. In: *Proceedings of the 41th international conference on Very Large Data Bases*
- Huang Z, Du W, Chen B (2005) Deriving private information from randomized data. In: *Proceedings of ACM SIGMOD international conference on management of data*, pp 37–48
- IBM. Big data and analytics. <http://www-01.ibm.com/software/data/bigdata>
- Islam MZ, Brankovic L (2011) Privacy preserving data mining: a noise addition framework using a novel clustering technique. *Knowledge-based Syst* 24: 1214–1223
- Jagannathan G, Pillaipakkamnatt K, Wright RN (2012) A practical differentially private random decision tree classifier. *IEEE Trans Data Privacy* 5(1): 273–295
- Kambatla K, Kollias G, Kumar V, Grama A (2014) Trends in big data analytics. *J Parallel Distrib Comput* 74(7): 2561–2573
- Laney D (2010) 3-D data management: controlling data volume, velocity and variety. META Group Research Note
- LeFevre K, DeWitt DJ, Ramakrishnan R (2006) Workload-aware anonymization. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp 277–286
- Li Y, Shen H (2010) On identity disclosure in weighted graphs. In: *Proceedings of the 11th international conference on parallel and distributed computing, applications and technologies*, pp 166–174
- Li N, Li T, Venkatasubramanian S (2007) t-closeness: Privacy beyond k-anonymity and l-diversity. In: *Proceedings of IEEE 23rd international conference on data engineering (ICDE)*, pp 106–115
- Li N, Qardji W, Su D, Cao J (2012) Privbasis: frequent itemset mining with differential privacy. *Proc VLDB Endowment* 5(11):1340–1351
- Liu K, Terzi E (2008) Towards identity anonymization on graphs. In: *Proceedings of ACM SIGMOD international conference on management of data*, pp 93–106
- Liu JQ, Wang K (2010) Anonymizing transaction data by integrating suppression and generalization. In: *Proceedings of the 14th Pacific-Asia conference on advances in knowledge discovery and data mining (PAKDD)*, pp 171–180
- Liu X, Yang X (2011) A generalization based approach for anonymizing weighted social network graphs. In: *Proceedings of the 12th international conference on web-age information management*, pp 118–130
- Liu L, Wang J, Liu J, Zhang J (2009) Privacy preservation in social networks with sensitive edge weights. In: *Proceedings of the SIAM international conference on data mining*, pp 954–965
- Liu L, Liu J, Zhang J (2010) Privacy preservation of affinities in social networks. In: *Proceedings of international conference on information systems*
- Machanavajjhala A, Kifer D, Gehrke J, Venkatasubramanian M (2007) l-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Discov Data* 1: Article 3
- Matwin S (2013) Privacy-preserving data mining techniques: survey and challenges. In: *Discrimination and privacy in the information society*. Springer, pp 209–221
- McAfee A, Brynjolfsson E (2012) Big data: the management revolution. *Harv Bus Rev* 90:60–68
- Meyerson A, Williams R (2004) On the complexity of optimal k-anonymity. In: *Proceedings of the 23rd symposium on principles of database systems (PODS)*, pp 223–228
- Mohammed N, Chen R, Fung BC, Yu PS (2011) Differentially private data release for data mining. In: *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, pp 493–501
- Mohan P, Thakurta A, Shi E, Song D, Culler D (2012) Gupt: privacy preserving data analysis made easy. In: *Proceedings of the 2012 ACM SIGMOD international conference on management of data*, pp 349–360
- Monreale A, Rinzivillo S, Pratesi F, Giannotti F, Pedreschi D (2014) Privacy-by-design in big data analytics and social mining. *EPJ Data Sci* 3:1
- Motwani R, Narab SU (2008) Anonymizing unstructured data. arXiv: 0810.5582v2, [cs.DB]
- Munir AB, Yasin SHM, Muhammad-Sukki F (2015) Big data: big challenges to privacy and data protection. *Int J Soc Behav Educ Econ Manag Eng* 9(1):355–363
- Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: *Proceedings of IEEE symposium on security and privacy*, pp 111–125
- Ni W, Chong Z (2012) Clustering-oriented privacy-preserving data publishing. *Knowledge-based Syst* 35:264–270
- Open, Public, Electronic, and Necessary Government Data Act of January 2019. <https://www.congress.gov/bill/115th-congress/senate-bill/760>
- Park H, Shim K (2007) Approximate algorithms for k-anonymity. In: *Proceedings of the 2007 ACM SIGMOD international conference on management of data*, pp 67–78
- Pearson S, Benameur A (2010) Privacy, security and trust issues arising from cloud computing. In: *Proceedings*

- of the 2nd IEEE international conference on cloud computing technology and science, pp 693–702
- Philip Chen CL, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Inf Sci* 275(10):314–347
- Punagin S, Arya A (2017) Privacy in the age of pervasive Internet and big data analytics-challenges and opportunities. *Int J Modern Educ Comput Sci* 7:36–47
- Ransing RS, Patole MS (2014) Data anonymization using map reduce on cloud by using scalable two-phase top down specialization approach. *Int J Sci Res* 3(12): 1916–1919
- Roja P, Subalakshmi R, Suganthi A, Sankaran L (2015) MapReduce concept for privacy preservation and data anonymization. *Discovery* 31(138):32–40
- Sagiroglu S, Sinanc D (2013) Big data: a review. In: Proceedings of international conference on collaboration technologies and systems, pp 42–47
- Samarati P (2001) Protecting respondents' identities in microdata release. *IEEE Trans Knowl Data Eng* 13(6):1010–1027
- Samarati P, Sweeney L (1998) Generalizing data to provide anonymity when disclosing information. In: Proceedings of ACM symposium on principles of database systems, p 188
- Savitz E (2012) Gartner: top 10 strategic technology trends for 2013. <http://www.gartner.com/newsroom/id/2209615>
- Serjantov A, Danezis G (2002) Towards an information theoretic metric for anonymity. In: Privacy enhancing technologies, pp 41–53
- Sweeney L (2002) K-anonymity: a model for protecting privacy. *Int J Uncertainty Fuzziness Knowledge-based Syst* 10(5):557–570
- Terrovitis M, Mamoulis N, Kalnis P (2008) Privacy-preserving anonymization of set-valued data. In: Proceedings of the international conference on very large databases (VLDB), pp 115–125
- Terrovitis M, Mamoulis N, Kalnis P (2011) Local and global recoding methods for anonymizing set-valued data. *VLDB J* 20(1):83–106
- The New York Times (2012) How companies learn your secrets. <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>
- United Nations Global Pulse (2012) Big Data for development: opportunities and challenges. <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>
- Vennila S, Pryadarshini J (2015) Scalable privacy preservation in big data a survey. *Procedia Comput Sci* 50: 369–373
- Wang SL, Tsai YC, Kao HY, Hong TP (2010) Anonymizing set-valued social data. In: Proceedings of the IEEE international symposium on social computing and networking (SocialNet), pp 809–812
- Wang SL, Tsai YC, Kao HY, Hong TP (2011a) Extending suppression for anonymization on set-valued data. *Int J Innov Comput Inf Control* 7(12):6849–6863
- Wang SL, Tsai YC, Kuo HY, Hong TP (2011b) K-anonymity on sensitive transaction items. In: Proceedings of the IEEE international conference on granular computing (GrC), pp 723–727
- Weiss R, Zgorski LJ (2012) Obama administration unveils “Big Data” initiative: announces \$200 million in new R&D investments. Office of Science and Technology Policy Executive Office of the President
- Wu X, Zhu X, Wu GQ, Ding W (2013) Data mining with big data. *IEEE Trans Knowl Data Eng* 26(1):97–107
- Xiao X, Tao Y (2006) Anatomy: simple and effective privacy preservation. In: Proceeding of the international conference on Very Large Databases (VLDB), pp 139–150
- Xu J, Wang W, Pei J, Wang X, Shi B, Fu AW-C (2006) Utility-based anonymization for privacy preservation with less information loss. *ACM SIGKDD Explorations Newsletter* 8(2):21–30
- Xu L, Jiang C, Wang J, Yuan J, Ren Y (2014) Information security in big data: privacy and data mining. *IEEE Access* 2:1149–1176
- Yang Y, Zhang Z, Miklau G, Winslett M, Xiao X (2012) Differential privacy in data publication and analysis. In: Proceedings of the 2012 ACM SIGMOD international conference on management of data, pp 601–606
- Yin S, Kaynak O (2015) Big data for modern industry: challenges and trends. In: Proceeding of IEEE, pp 143–146
- Yuan M, Chen L, Yu PS (2010) Personalized privacy protection in social networks. In: Proceedings of the 36rd international conference on Very Large Data Bases, pp 141–150
- Zaherzadeh H, Aggarwal CC, Barker K (2015) Privacy-preserving big data publishing. In: Proceedings of the 27th international conference on scientific and statistical database management
- Zhang X, Liu C, Nepal S, Yang C, Chen J (2014a) Privacy preservation over big data in cloud systems. In: Security, privacy and trust in cloud systems. Springer, pp 239–257
- Zhang X, Dou W, Pei J, Yang C, Liu C, Chen J (2014b) Proximity-aware local-recoding anonymization with MapReduce for scalable big data privacy preservation in cloud. *IEEE Trans Comput* 64(8):2293–2307
- Zhang X, Yang LT, Liu C, Chen J (2014c) A scalable two-phase top-down specialization approach for data anonymization using MapReduce on cloud. *IEEE Trans Parallel Distrib Syst* 25(2):363–373
- Zhou B, Pei J (2008) Preserving privacy in social networks against neighborhood attacks. In: Proceedings of the 24th international conference on data engineering, pp 506–515
- Zikopoulos P, Eaton C, Zikopoulos P (2011) Understanding big data: analytics for enterprise class hadoop and streaming data. McGraw Hill Professional
- Zou L, Chen L, Ozsu MT (2009) K-automorphism: a general framework for privacy preserving network publication. In: Proceedings of the 35rd international conference on Very Large Data Bases, pp 946–957

Access Control for XML Big Data Applications

Alberto De la Rosa Algarin¹,
Steven A. Demurjian² and Eric Jackson³

¹Cynnovative, LLC, Arlington, VA, USA

²Department of Computer Science and
Engineering, University of Connecticut, Storrs,
CT, USA

³Department of Civil and Environmental
Engineering, University of Connecticut, Storrs,
CT, USA

Article Outline

Introduction
Use-Case Background
UML Extensions for Document Modeling
Extensions for Document Security Policy
 Modeling and Integration
Integrating Local Security Policies into a Global
 Security Policy
Related Work
Conclusion
Bibliography

Introduction

Big data is defined as a large volume of data, both structured and unstructured, which engulfs a business on a day-to-day basis (<https://www.sas.com/en-us/insights/big-data/what-is-big-data.html>). As a concept, “big data” has an implicit association to higher storage and processing requirements, complex security, and costly maintenance. Domains such as law enforcement, health care, e-commerce, and national defense have all adopted the term “big data” to describe vital parts of their daily workflow and complex systems. Due to the sensitivity of the data these domains work with, security assurance has emerged as an important requirement to meet

when developing applications that make use of it to achieve a specific goal. Typically, big data applications are built as a system-of-systems. That is, these applications aim to fuse separate repositories of data in order to provide a common operating picture to the end user. To achieve this, big data applications share and exchange a wide variety and high volume of information in potentially different formats (e.g., XML, RDF, JSON, etc.). In turn, this creates conflicting security capabilities of constituent systems (the components of the big data application), the diverse set of stakeholders (the owners of the big data stores), and the overall security logic to be attained. Any security approach that aims to control the mechanisms of access to the underlying data must support granularity to the point of “how,” “what,” and “when” stakeholders can access. In addition, security policies of constituent systems must be met, assured, and reconciled with one another to produce a global security policy that is consistent and usable.

In this entry, we consider access control for big data applications using three dominant access control models: mandatory access control (MAC) (Bell and La Padula 1976) that assigns sensitivity levels to subjects (clearances) and objects (classifications) to control access to information; role-based access control (RBAC) (Ferraiolo et al. 2001) that focuses on the responsibilities of the users for an application per role; and discretionary access control (DAC) (Sandhu and Samarati 1994). Mandatory access control utilizes security sensitivity levels that are assigned to subjects (clearance) and objects (classification) with the permissions for the subject to read and/or write an object dependent on the relationship between clearance and classifications. MAC typically is modeling using four sensitivity levels, which are hierarchically ordered from most to least secure: top secret (TS) < secret (S) < classified (C) < unclassified (U). Role-based access control (RBAC) allows the responsibilities of the role within the application to dictate the permissions for a given user. Roles are defined in an application in order to name entities

that are often related to job categories and responsibilities that can be extrapolated from individual users and established as a reusable role. In discretionary access control (DAC), access to objects is permitted or denied based on the subject's identity or the group they are part of with the ability to have privileges pass among users either under the control of a user or as dictated by a security administrator. DAC utilizes the concept of delegation to be able to pass privileges among users in certain situations and under specific constraints.

In this entry, we will present and explain a framework to represent, integrate, and reconcile security policies of constituent systems in support of access control (MAC, RBAC, and DAC) for a big data application that is modeled using eXtensible Markup Language (XML) (<https://www.w3.org/XML/>). The work presented in this entry leverages our prior work that included a security model that we have implemented (De la Rosa Algarín and Demurjian 2013, 2016; De la Rosa Algarín et al. 2012, 2013a, b; De la Rosa Algarín 2014) based on a generalized tree form, specifically, supporting a document model where each document has a schema and each schema has a single root node and potentially many children nodes. Second, a majority of the RBAC capabilities in the NIST model are supported through the definition of roles, the assignment of operations as permissions, and the determination of constraints in support of role delegation with mutual exclusion. Third, a wide range of MAC capabilities will be supported by the security model, allowing a broad range of information systems and their security needs to be supported including the ability to assign classifications to all application schemas, their elements, and define clearances for users. This assignment extends the schema via a decoration operation that acts over a determined criterion and results in an extension of the schema and its elements (e.g., the entire document tree) with classifications as tags that are then enforced at the instance level for actual users holding the appropriate clearance. When RBAC and MAC are combined, a given role can access each element in a specific way via a particular operation, achieved by filtering schemas and instances utilizing the role as the criterion. Fourth,

the ability to support DAC that includes the delegation of role from user to user and the ability to pass on the delegation. The model completes with a discussion from the user perspective that includes authorizations over schemas and instances. Finally, for the purposes of this entry, we note the distinction between document-level security and policy-level security. *Document-level security* includes those concepts that act on the tree-structure (e.g., classification levels for elements, access modes of operations that target the elements, operations and permissions for a role that are tied to an element, etc.), while *policy-level security* pertains to those capabilities that are defined at a higher level (e.g., authorizations over schemas and instances). The main assertion toward the research in this entry is that RBAC and MAC are orthogonal. That is, their capabilities do not conflict with one another. This is achieved by assigning clearance levels to users and not roles. Ultimately, when unifying RBAC, MAC, and DAC, MAC requirements take a higher-level priority and dictate the final security effect; but from a construction perspective, the security can be added in any order with the end result the same in terms of the defined and enforced permissions. For the purpose of this entry, we specialize the document-level Security to apply to XML documents and the associated instances that are part of the big data application.

The work presented in this entry extends our previous work (De la Rosa Algarín et al. 2014) in which we focused on solely RBAC concerns in XML represented documents, to focus on a combination of Mandatory, Role-Based, and Discretionary Access Control (MAC (Bell and La Padula 1976), RBAC (Ferraiolo et al. 2001), and DAC (Sandhu and Samarati 1994), respectively) assurance from the point of the big data application that uses any type of tree-structured document format (e.g., XML, JSON, RDF, etc.). To demonstrate the feasibility of this approach, we leverage a use-case in the law enforcement domain, specifically traffic accident and crash-related data. We demonstrate that this approach can generate a global policy for the big data application that targets the underlying data provided by each constituent system. The resulting global

security policy is heavily structured around data sharing from constituent systems. In this entry, in addition to supporting a combination of MAC, RBAC, and DAC, we specialize on our prior work that was focused on document-level RBAC to support all three kinds of access control models for the XML schemas and associated document instances that are part of big data applications.

The remainder of this entry has six sections. Section “[Use-Case Background](#)” provides a brief background on the use-case that support the examples in this entry, specifically the Connecticut Crash Data Repository (CCDR) on motor vehicle crashes on all state and interstate roads recorded by local and state law enforcement authorities in the state of Connecticut to demonstrate the work in this entry. Section “[UML Extensions for Document Modeling](#)” describes our framework capabilities to represent XML using UML. Specifically, we explain and demonstrate the way that the Document Schema Class Diagram (DSCD) can model the information for a big data application. Section “[Extensions for Document Security Policy Modeling and Integration](#)” provides an overview of the MAC, RBAC, and DAC UML extensions in our framework. Section “[Extensions for Document Security Policy Modeling and Integration](#)” also applies this framework with policy modeling and integration capabilities to capture local and global security policies in support of the security for a big data application. Section “[Integrating Local Security Policies into a Global Security Policy](#)” details the process of integrating multiple local security policies into a global policy for a big data application, including reconciling conflicts across local systems, roles, and permissions in order to attain a global policy that can be assured. Section “[Related Work](#)” details related work on policy/security modeling and policy integration, with section “[Conclusion](#)” offering concluding remarks.

Use-Case Background

United States law enforcement roles and tasks are organized in a hierarchical fashion. Cities and

counties have their own police forces, with an overarching state police force. Road management, for example, also follows a hierarchical structure with local police concentrating on their local roads and the state police managing state and interstate highways. Federal components, such as the Federal Bureau of Investigation (FBI) focus at the national level and when crimes cross state boundaries. Every law enforcement officer, regardless of their rank, generates data regarding traffic accidents, criminal activities, and prevention, etc. As a result, for any one particular law enforcement case, there is a potential for a wide range of data to be available and linked within and across different states. Perfect examples of such big data system are the Integrated Automated Fingerprint Identification System (IAFIS) and the National Missing and Unidentified Persons System (NamUs) (<https://www.fbi.gov/services/information-management/foipa/privacy-impact-assessments/iafis>; <https://www.namus.gov/>).

Another example of such a repository is the Connecticut Crash Data Repository (CCDR) collection (Fig. 1) on motor vehicle crashes by law enforcement in the state of Connecticut that will be utilized to demonstrate the work in this entry. This CCDR has a large amount of stakeholders, including the Civil Engineering and Computer Science and Engineering departments at the University of Connecticut as the leaders in its development, the Connecticut Department of Transportation (DoT), the Connecticut Police Department (and all regional departments), researchers that request access to the data in support of studies, and even lawyers accessing data for liability research. Other states have similar repositories of data, including Florida, Indiana, Wisconsin, Louisiana, and New Jersey (<http://www.flhsmv.gov/ddl/ecrash/>; <http://www.in.gov/cji/2481.htm>; <http://transportal.cee.wisc.edu/services/crash-data/>; <http://datareports.lsu.edu/>; <http://www.state.nj.us/transportation/refdata/accident/>).

From a security perspective, there are a wide variety of workflows and scenarios where the data from these collections can be leveraged. For instance, a DoT employee at the local, state, or federal level would be able to read and curate (correct) the crash data, a police officer would

Connecticut Crash Data Repository

Report Date Settings

From Year (YYYY): To Year (YYYY):

Available Reports

Select Reports:

Road Classification

Rural or Urban: Route Class:

Select Area

Town: Planning Agency: County: DOT District:

Optional Descriptions

Please Upload Report Logo: No file chosen Report Subtitle:

Connecticut Crash Data Query Tool

Query Result

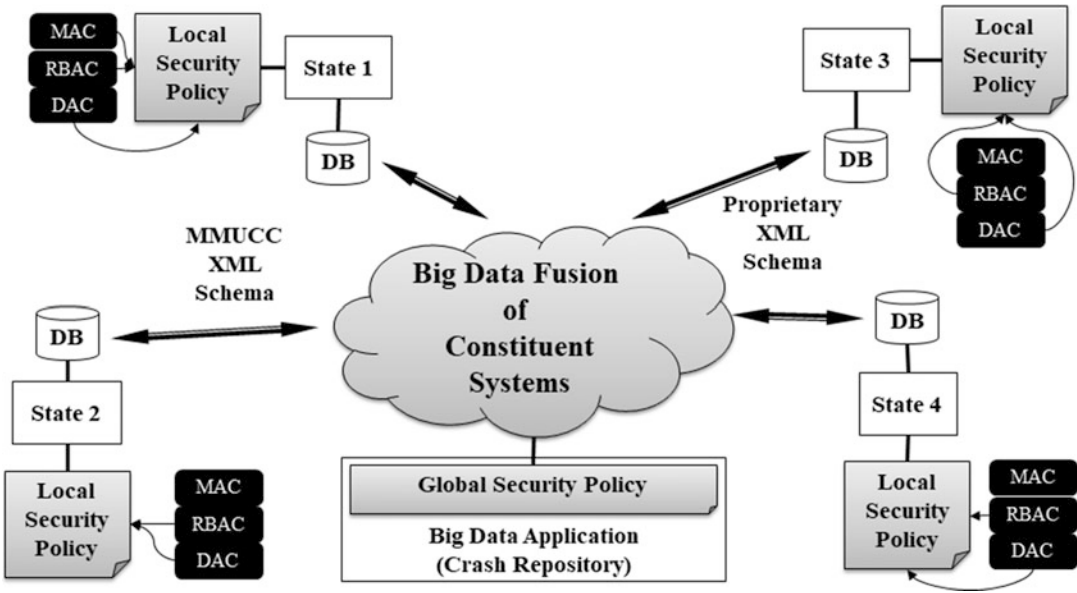
Records 1 - 25 of 125499

Record	Case Number	Crash Date	Crash Time	Crash Severity	Town	Route Class	Route Number	Cumulative Route Mileage	Road Name	Intersection	Crash Occurrence	Crash Type	Weather Condition	Road Surface Condition	Light Condition	Roadway Feature	Median Barrier Penetration	Construction Related	At Fault	Contributing Factor
1	300000	1997-01-01	13:24:00	2	144	3	131	1.80	CONN ROUTE 111	FOR TO GOLF GAS STATION	9	1	1	1	1	4	4	2	2	8
2	300001	1997-01-01	14:02:00	3	18	2	202	12.81	U.S. ROUTE 202	US 7 NB	9	1	1	1	1	4	2	2	5	
3	300002	1997-01-01	13:42:00	2	144	3	130	2.60	CONN ROUTE 130	ACCESS TO RTE 8 (RICHES)	1	1	1	1	1	4	2	2	3	
4	300003	1997-01-01	00:39:00	3	33	3	372	14.33	CONN ROUTE 372	HELLO POLE 120	5	2	2	3	0	4	2	2	10	
5	300004	1997-01-01	14:05:00	2	148	4	226	1.36	CONN ROUTE 226	NORTH ASHLIN RD	5	1	2	2	1	0	4	2	1	4
6	300005	1997-01-01	00:01:00	3	90	1	91	0.40	INTERSTATE 90	DE EXIT TO I-95 WEST	15	1	1	1	1	0	4	2	1	2
7	300007	1997-01-01	04:16:00	2	138	1	95	13.73	INTERSTATE 95	INTERSTATE 95	15	2	2	3	5	3	3	1	12	
8	300008	1997-01-01	00:36:00	3	144	3	13	30.67	CONN ROUTE 13	NB ACC R RTE 111 & SR 72	9	1	2	1	8	3	2	2	8	
9	300010	1997-01-01	22:14:00	2	144	4	470	0.20	TELLER RD		15	1	4	2	0	4	2	1	9	
10	300011	1997-01-01	09:15:00	2	120	3	372	1.80	CONN ROUTE 372	RTE 372 - EAST MAIN ST	9	1	2	1	1	4	2	2	8	
11	300012	1997-01-01	01:52:00	3	88	3	68	0.86	CONN ROUTE 68	UNION HILL ST	3	2	2	3	1	4	2	1	5	
12	300013	1997-01-01							CONN ROUTE											

Access Control for XML Big Data Applications, Fig. 1 CTCDR Query Selection and Result Screens

enter new crash reports, a researcher would be interested in aggregate and other trend-oriented queries, or an attorney representing a client in an accident and wants to see any prior accidents at

that location (akin to an FOI request). What is not supported is the ability to fuse information across all crash data for all states (constituent systems), to form a big data application that will allow



Access Control for XML Big Data Applications, Fig. 2 Fusion of Big Data DaaS instances and Global Security Policy generation for FDR

queries that cross state boundaries. This entry demonstrates the way that our security framework provides security assurance and facilitates the development of a would-be big data application, which we will refer to as a Federated Crash Repository (FCR).

Figure 2 illustrates the big data application use-case (FCR), the underlying concepts, policy definition, and integration processes to link the local security policies (states) into a global context for the crash data (country). A big data application that uses the MMUCC schema to represent the data from the constituent systems’ DaaS is created. Each of the constituent systems contains potentially disparate local security requirements (Local Security Policy objects at each corner), which are then integrated via the framework presented in this entry. To achieve the Global Security Policy, we execute three major steps. First, we model the data schema (in this case, the MMUCC). Second, we model the local security requirements (each state’s local security policy), capturing all the RBAC, MAC, and DAC requisites. Last, we integrate the modeled security policies.

UML Extensions for Document Modeling

The unified modeling language, UML, provides a large variety of diagrams for the visualization of different software requirements: class, component, deployment, activity, use-case, state-machine, communication, sequence, etc. (Fowler 2004). Our prior work (De la Rosa Algarín and Demurjian 2013, 2016; De la Rosa Algarín et al. 2012, 2013a, b; De la Rosa Algarín 2014) introduced new UML diagrams that correspond to the key security model characteristic. First, there exists a necessity to represent a tree-structured schema. To this purpose, we use the *UML Document Schema Class Diagram (DSCD)*, which is an artifact that holds all of the characteristics of the data to be secured, including structure, data type, and value constraints. The DSCD graphically represents the schemas of data provided by a constituent system in the big data application scenario. To enable this process, we capitulate on the *UML Profile* paradigm, which allows new diagrams to be defined using the various UML concepts (stereotypes, tags, constraints applied to classes, attributes, operations, etc.). This permits

Access Control for XML Big Data Applications, Table 1 Specialized UML Profile for Tree-Structured Document to DSCD with XML Cases

Tree-structured document component	XML	DSCD component
General Element	Element	UML Class
General Element Name	Element Name	UML Class Name
General Element Attribute	Generic Attribute	Stereotyped Attribute
Parent–child relationship of a schema (tree-subtree)	Tree – Subtree	UML Dependency Relationship
Complex Type of Elements and/or Attributes	XML <i>xsd:complexType</i>	Stereotyped «complexType» UML Class
Sequential Element Order	XML <i>xsd:sequence</i>	Stereotyped «sequence» UML Class
Aggregation of Attributes	XML <i>xsd:attributeGroup</i>	Stereotyped «attributeGroup» UML Class
Grouping of Elements to form a complex type	XML <i>xsd:group</i>	Stereotyped «group» UML Class
Acceptable Values for Elements	XML constraints via <i>minOccurs</i> , <i>maxOccurs</i>	Stereotyped «constraint» Class Member
Indirect References of Elements	XML <i>ref</i>	Stereotyped «ref» Name Class Member
Parent–child relationship of non-named Elements	XML Element – non-named child element	UML Directed Associate Relationship

the representation of constituent system data, in the form of tree-structure documents, to be transformed into a DSCD and back (roundtrip engineering). Table 1 shows the way that the different components of a document's structure, such as those represented in XML, have a matching DSCD component in our framework's UML profile.

We chose to represent each document schema as a tree of *stereotyped classes*. This approach allows us to capture the hierarchical structure of a document as a series of related *classes*. Table 1 has three columns: the first column represents the features of a generic tree-structured document, the second column has the XML equivalent as an example, and the third column transforms the first two into the equivalent UML profile component. In the first row of Table 1, a general element in the tree-structured document is equivalent to an XML element (*xsd:element*) and is realized as a UML class; the second row maps the element name to a UML class name. In the third row of Table 1, an element attribute in the tree-structured document is equivalent to a generic attribute in XML, which can be mapped to a «stereotyped» attribute in UML. The fourth row corresponds to a

patient–child relationship at the schema level to identify a tree and its subtrees, which in XML is observed as nested elements, and is represented as a UML dependency relationship in the DSCD. The fifth row of Table 1 describes complex elements (those that are built out of many sub-elements), which in XML are denoted as *xsd:complexType* and in the DSCD are denoted as a UML class with the «complexType» stereotype. The sixth row covers a similar case, considering sequences or lists of elements, which in XML are denoted as *xsd:sequence* and in the DSCD are denoted as a UML class with the «sequence» stereotype. Aggregation of attributes are handled with the seventh row of Table 1 and is represented as *xsd:attributeGroup* in XML and as a UML class with the «attributeGroup» stereotype in the DSCD. In the eighth row of Table 1, groups of elements in a tree-structured document are equivalent to an XML *xsd:group* node and is represented as a UML class with the «group» stereotype in DSCD. The ninth row of Table 1 handles acceptable or allowable values for elements, which in XML are usually *maxOccurs* and *minOccurs* attributes to an XML element constraints, realized as a «constraint» stereotyped

class member in DSCD. In the tenth row of Table 1, indirect references allow elements of a tree-structure document to be associated with one another, which in XML is a ref. attribute on an element that are represented as a «ref.» class member from UML profile in the DSCD. Lastly, in the eleventh row of Table 1, for tree-structured document, the parent–child relationship between non-named elements corresponds to non-named elements in XML (e.g., *xsd:complexType*, *xsd:attributeGroup*, etc.) and is represented with a UML directed association relationship between classes in the DSCD.

As an example of utilizing the UML profile approach to convert a segment of data into a DSCD, let us use the MMUCC schema (Guideline 2012) from the law enforcement domain to represent motor vehicle crashes. The conversion of the schema from Fig. 3 into a DSCD follows a mapping process guided by the profile of Table 1, resulting in a DSCD as shown in Fig. 4.

Extensions for Document Security Policy Modeling and Integration

To supplement the DSCD of section “UML Extensions for Document Modeling” in support of demonstrating access control for big Data applications like FCR, we review a number of other diagrams (De la Rosa Algarín 2014). The *Secure Information Diagram (SID)* identifies the subset of DSCD that needs to be protected. For elements of a schema, permissions for RBAC (read, write, etc.), and classifications for MAC, we utilize UML *Document Role-Slice Diagram (DRSD)*. To support MAC feature, the UML *MAC Secure Information Diagram (MSID)* builds off of the SID and serves as a means to define which elements in the tree-structured document need classifications. To tie together RBAC, MAC, and DAC, the *User Diagram (UD)* considers the orthogonal nature between RBAC and MAC, as well as the role-delegation and authorization assignments from DAC. To support the ability to have privileges delegated from one user to another, the UML *Delegation Diagram (DD)*

provides which users and role pairs can delegate or be delegated. Lastly, to support the definition of users and their authorization, the UML *Authorization Diagram (AD)* provides the ability to determine which schemas and instances are tied to a user/role pair. These UML model extensions to represent MAC, RBAC, and DAC security requirements are accomplished by extending the UML Meta-Object Facility (MOF, Fig. 5), which enables extensions to the modeling language to be defined with several degrees of formality.

The document schemas, as shown in the MMUCC example DSCD in Fig. 4, can be constrained to identify those portions of the schema that require security control. To achieve this, the *Secure Information Diagram (SID)* is used to represent those portions (elements and subtrees) of an application’s schema, which both MAC classifications and RBAC permissions need to be defined. For the SID, the M2 metamodel is shown in Fig. 6, where each class that is part of the SID is represented as meta-class (*SecureInformation*) associated with many possible instances of any given schema element as represented with the *Element* meta-class. Following the example of the MMUCC, the realization of the SID is shown in Fig. 7, a subset of the information from Fig. 4 that needs to be secure.

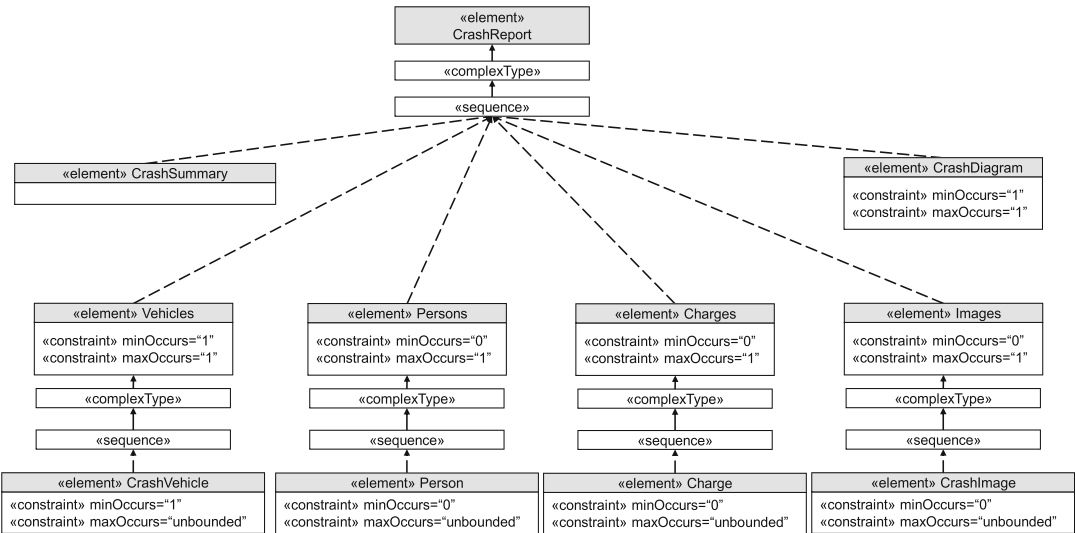
To support RBAC role hierarchy, RBAC operations that target schemas and their instances, and granular MAC classification of elements, we use the *Document Role-Slice Diagram (DRSD)*. The metamodel for the DRSD (as shown in Fig. 8) consists of the *RoleSlice* meta-class, which represents the role slices that will be defined with permissions against the SID (see Fig. 7 again) with respect to the schema(s) of the application to be secured. The *Permission* meta-class represents the permissions allowed over the instances validated against the secured schema (read, aggregate, insert, update, delete) that define what a role can and can’t do for the elements in a schema. In order to create a relation between the *RoleSlice* meta-class (which contains all of the DRSD instances) and the *Permission* meta-class (which contains all of the schema targeting permissions), it is necessary to create a relation between the users and their roles. In Fig. 7, the *UserRole*

```

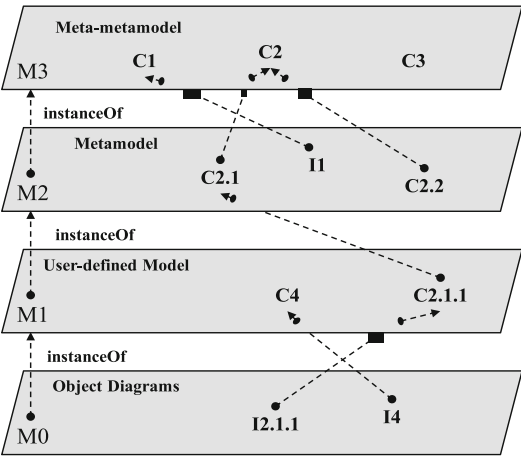
1.  <?xml version="1.0" encoding="utf-16"?>
2.  <xsd:schema
3.    xmlns:b=http://schemas.microsoft.com/BizTalk/2003
4.    xmlns=http://CDR.CollisionReport/V2/0
5.    elementFormDefault="qualified"
6.    targetNamespace=http://CDR.CollisionReport/V2/0
7.    xmlns:xsd="http://www.w3.org/2001/XMLSchema">
8.    <xsd:include
9.      schemaLocation=".\\crashreportbase.xsd" />
10.    <xsd:annotation>
11.      <xsd:appinfo>
12.        <b:schemaInfo root_reference="CrashReport"
13.          xmlns:b="http://schemas.microsoft.com/BizTalk/2003" />
14.      </xsd:appinfo>
15.    </xsd:annotation>
16.    <xsd:element name="CrashReport">
17.      <xsd:complexType>
18.        <xsd:sequence>
19.          <xsd:element ref="CrashSummary" />
20.          <xsd:element minOccurs="1" maxOccurs="1" name="Vehicles">
21.            <xsd:complexType>
22.              <xsd:sequence>
23.                <xsd:element minOccurs="1" maxOccurs="unbounded"
24.                  ref="CrashVehicle" />
25.              </xsd:sequence>
26.            </xsd:complexType>
27.          </xsd:element>
28.          <xsd:element minOccurs="0" maxOccurs="1" name="Persons">
29.            <xsd:complexType>
30.              <xsd:sequence>
31.                <xsd:element minOccurs="0" maxOccurs="unbounded"
32.                  ref="Person" />
33.              </xsd:sequence>
34.            </xsd:complexType>
35.          </xsd:element>
36.          <xsd:element minOccurs="0" maxOccurs="1" name="Charges">
37.            <xsd:complexType>
38.              <xsd:sequence>
39.                <xsd:element minOccurs="0" maxOccurs="unbounded"
40.                  ref="Charge" />
41.              </xsd:sequence>
42.            </xsd:complexType>
43.          </xsd:element>
44.          <xsd:element minOccurs="1" maxOccurs="1"
45.            ref="CrashDiagram" />
46.          <xsd:element minOccurs="0" maxOccurs="1" name="Images">
47.            <xsd:complexType>
48.              <xsd:sequence>
49.                <xsd:element minOccurs="0" maxOccurs="unbounded"
50.                  ref="CrashImage" />
51.              </xsd:sequence>
52.            </xsd:complexType>
53.          </xsd:element>
54.        </xsd:sequence>
55.      </xsd:complexType>
56.    </xsd:element>
57.  </xsd:schema>

```

Access Control for XML Big Data Applications, Fig. 3 Segment of MMUCC Schema



Access Control for XML Big Data Applications, Fig. 4 A DSCD for the MMUCC Segment



Access Control for XML Big Data Applications, Fig. 5 UML's Meta-Object Facility Layers

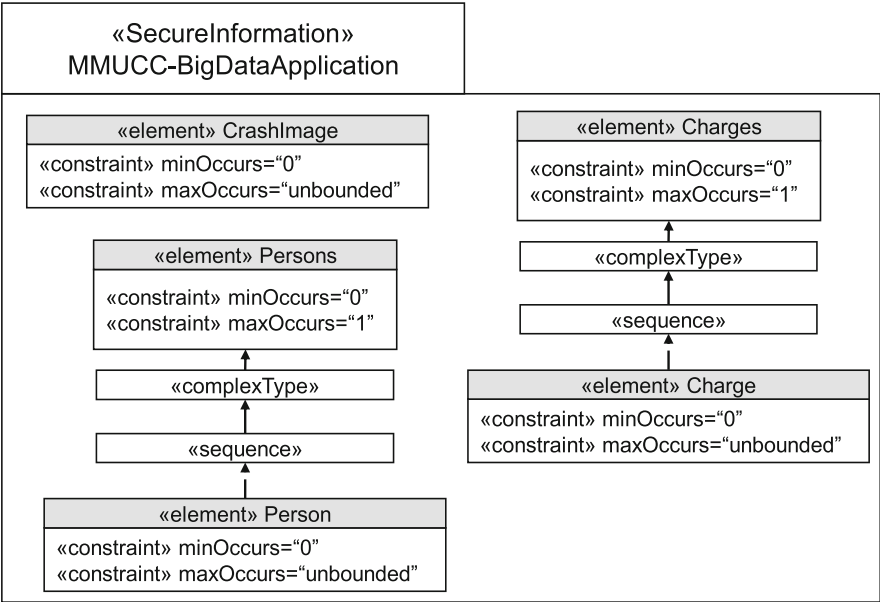


Access Control for XML Big Data Applications, Fig. 6 Secure Information Diagram M2 metamodel

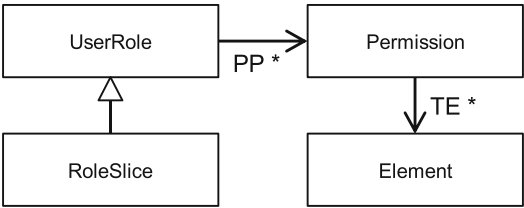
meta-class is a parent-class of the *RoleSlice* meta-class and a sibling class of the *Permission* meta-class. The connections between the *UserRole* and *Permission* meta-classes are given by the *permitted permission* (PP) relation. The *Element* meta-

class in turn represents all of the instances of elements (from the schema) that are targeted by the different permissions. This connection is tagged with the targeted element (TE) label in Fig. 8. Again, following the example of the MMUCC, the realization of the DRSD with sample roles against the SID of Fig. 7 is shown in Fig. 9.

To support MAC, we use the *MAC Secure Information Diagram* (MSID). As shown in Fig. 10, the MSID is a UML package with the stereotype «SecureInformation» that decorates the SID and contains all of the respective classes of elements from the schema to be secured per access modes (ams) and classifications (cls). The meta-model for the MSID consists of four meta-classes: *User*, *AccessMode*, *Element*, and *SensitivityLevel*. These metaclasses are interconnected to represent the relations between the user (*User*), its clearance level (*Sensitivity*), and access modes (read, aggregate, insert, update, delete; *AccessMode*) for each of the elements (xs: element, xs:complexType, etc.; *Element*) from the SID that need to be protected. To represent the relation between *User* and *SensitivityLevel*, an arrow with a UC (user clearance) tag is used. This relation indicates that the user could either have a clearance level or is without a clearance level, therefore the utilization of the 0..1



Access Control for XML Big Data Applications, Fig. 7 SID with MMUCC Elements



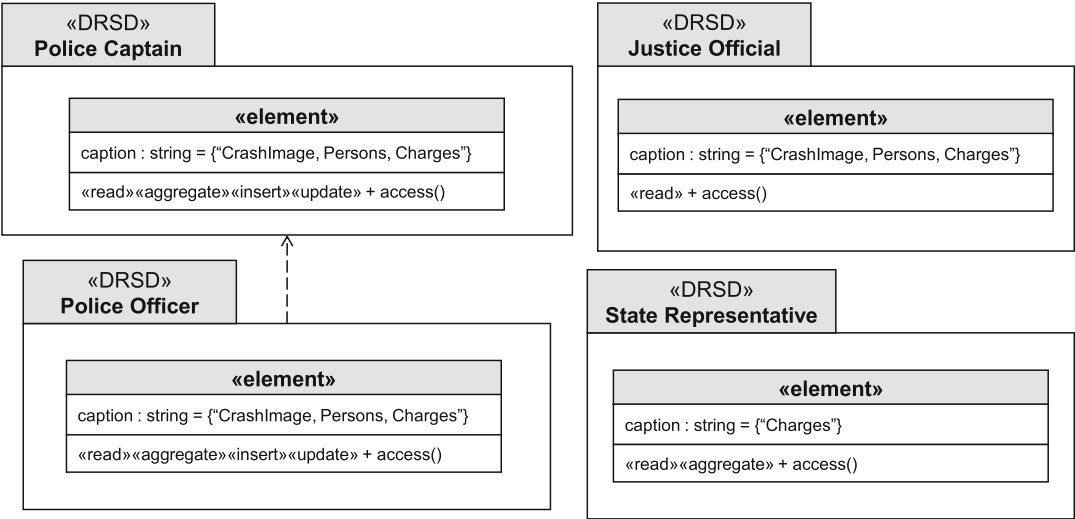
Access Control for XML Big Data Applications, Fig. 8 Document Role Slice Diagram M2 metamodel

cardinality constraint. Element and Sensitivity are similarly related, represented with the arrow tagged *EC* (element classification). The relationship between *Element* and *AccessMode* is represented with the *1..+* cardinality constraint to cover the case of an element with different possible access modes. The result of the meta-model instantiation with the MMUCC example is shown in Fig. 11.

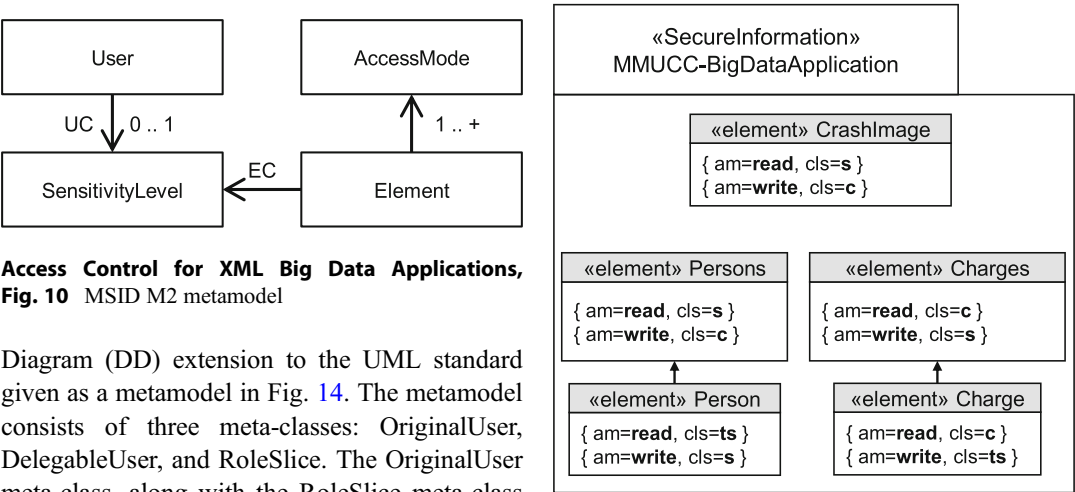
Next, at the metamodel layer (M2) of the MOF, the User Diagram (UD) extension is presented in Fig. 12. The interplay of users, their roles and delegation permissions (for RBAC), their clearance levels (for MAC), and their authorization permissions (DAC) require the proper definition of a user concept. The work in secure software engineering (Pavlich-Mariscal et al. 2010)

proposed a UML extension for users via a User Diagram. In this entry, we build upon this first iteration of the User Diagram to extend to include both LBAC and RBAC user features directly to the metamodel. The metamodel of the UD is composed of six major meta-classes as given in Fig. 11: User, SensitivityLevel, UserRole, RoleSlice, SOD, and ME. The User meta-class represents all of the possible instances of users in a particular application. Both the User meta-class and the RoleSlice meta-class is a subtype of the UserRole meta-class. The UR tag represents user-role assignments (RBAC), the separation of duty (SOD) meta-class represents the separation of duty relations, and the mutual exclusion (ME) meta-class represents the mutual exclusion relations between roles. The SensitivityLevel meta-class, which represents the sensitivity as related to LBAC is a clearance level for a user and is tied to the User meta-class. This distinction shows an important feature of the security framework presented in this dissertation, namely, that RBAC and LBAC capabilities are orthogonal. Following the examples of the roles in Fig. 9, we show an example UD for the MMUCC in Fig. 13.

The delegation component of the underlying security model is supported with the Delegation



Access Control for XML Big Data Applications, Fig. 9 DRSD Role Hierarchy for MMUCC Access in the Law Enforcement Scenario



Access Control for XML Big Data Applications, Fig. 10 MSID M2 metamodel

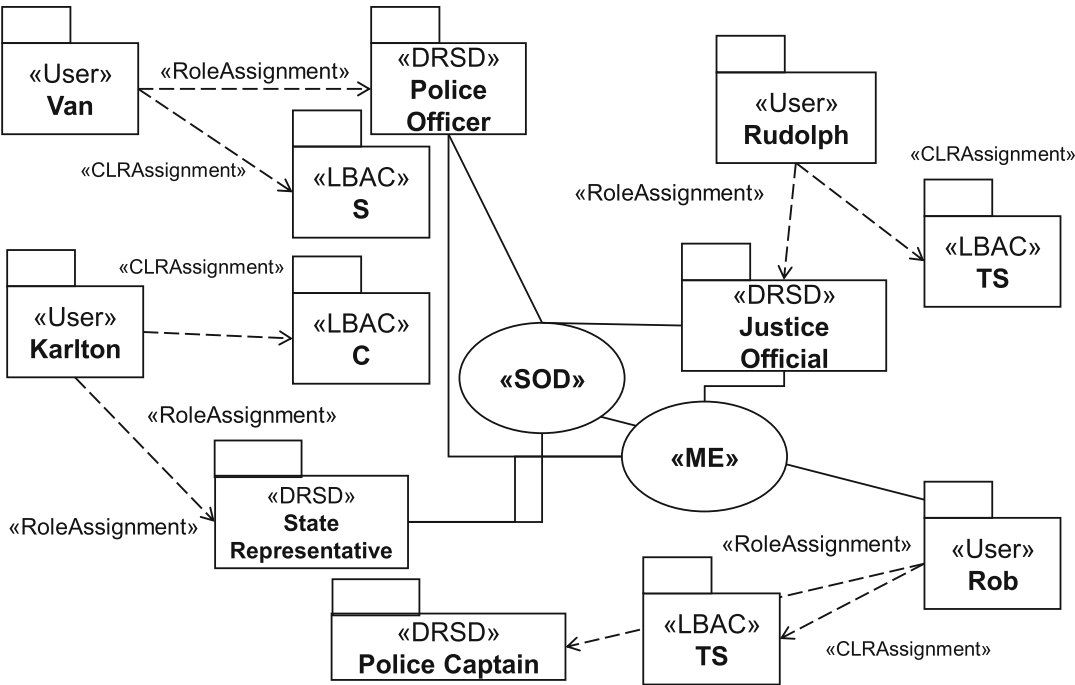
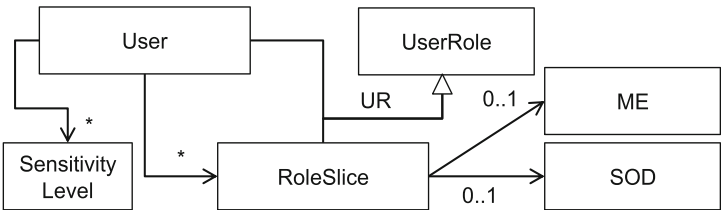
Diagram (DD) extension to the UML standard given as a metamodel in Fig. 14. The metamodel consists of three meta-classes: OriginalUser, DelegableUser, and RoleSlice. The OriginalUser meta-class, along with the RoleSlice meta-class represents the original users of the application and their assigned roles. The DelegableUser, connected to the RoleSlice meta-class, represents the user/role pairs of authorized delegations. In turn, the Delegation tag in the connection between OringalUsers and DelegableUsers represents the ability to perform the delegation operation. Figure 15 shows how a sample instantiation of the DD metamodel would look in the example of the MMUCC.

The final extension of the UML standard to support the security model consists of the Authorization Diagram (AD) metamodel in Fig. 16 that

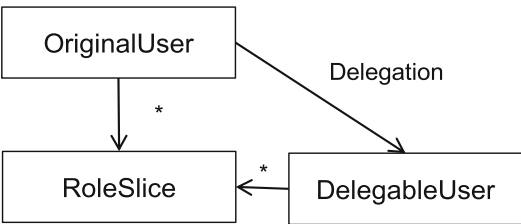
Access Control for XML Big Data Applications, Fig. 11 MSID for MMUCC Access

consists of four meta-classes: UserRole, Authorization, Instance, and Schema. The UserRole meta-class represents the specific user/role pair in a similar fashion as the case of the UD in Fig. 13. The Authorization meta-class is connected to the Instance and Schema meta-classes to represent whether an authorization to an instance or schema exists and is represented with the 0..+ tag on the directional connection. This metamodel definition allows scenarios in

Access Control for XML Big Data Applications, Fig. 12 User Diagram M2 metamodel



Access Control for XML Big Data Applications, Fig. 13 User Diagram for the Law Enforcement Scenario



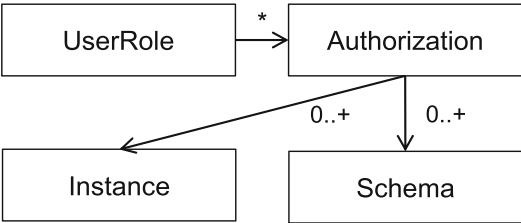
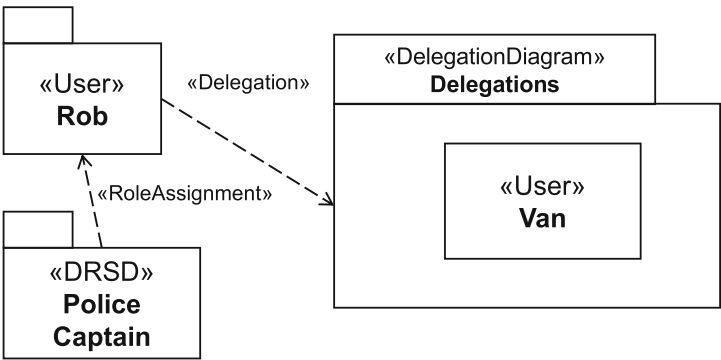
Access Control for XML Big Data Applications, Fig. 14 Role Delegation Diagram M2 metamodel

which a user might not be authorized to any schema/instance, or any combination of the two (e.g., all schemas and all instances). Figure 17 shows an MMUCC-related AD with the users and roles utilized from Figs. 13 and 9, respectively.

Integrating Local Security Policies into a Global Security Policy

With the constituent security policies modeled as local instances of DRSD, MSID, UD, and DD as shown in Fig. 18 and as discussed in section “Extensions for Document Security Policy Modeling and Integration,” policy integration can be performed at the model level to generate a Global Security Policy (GSP) for the big data application. Unlike our previous work, in which we integrate local security policies with RBAC concerns only, here we present a process to integrate a set of local instances of security policy concerns (e.g., for each constituent system or state in the law enforcement example) that cover

Access Control for XML Big Data Applications, Fig. 15 DD for User Rob in Police Captain Role in a Law Enforcement Scenario



Access Control for XML Big Data Applications, Fig. 16 Authorization Diagram M2 metamodel

MAC, RBAC, and DAC into a GSP for the new big data application. Toward this goal, section “Assumptions and Equivalence Finding” details our assumptions and various new concepts; Section “Integration Process for Local Security Policy Sets” examines the integration process of the local security policies into a GSP; Section “Resolving Conflicts of Integrated Security Rule Sets” examines the security rule conflict resolution process; and, section “Creating the Global Security Policy” presents the GSP that is generated for the big data application drive by the law enforcement FCR example.

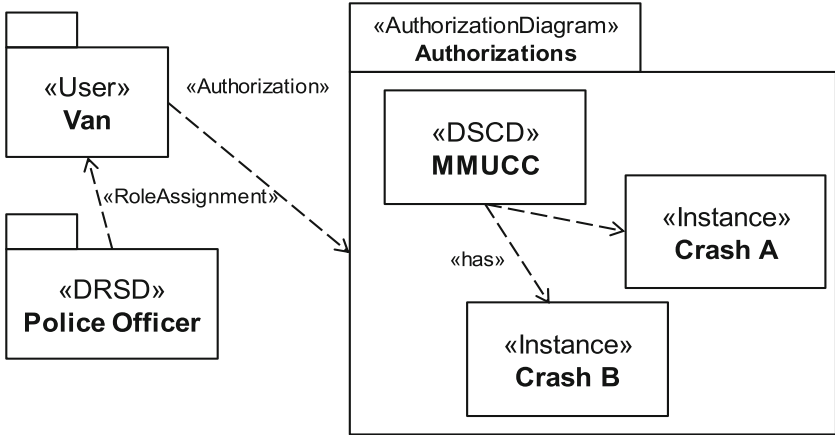
Assumptions and Equivalence Finding

We define a local security policy set as the combination of instances of diagrams from the previous sections (e.g., DSCD, SID, DRSD, MSID, UD, and DD) respective to a constituent system in a big data. The intent of the local security policy set is to provide the ability for each constituent local system (e.g., State 1 in Fig. 2) to define one or multiple security policies that govern the data they make available. The local security policy sets can be published for inclusion in the GSP that

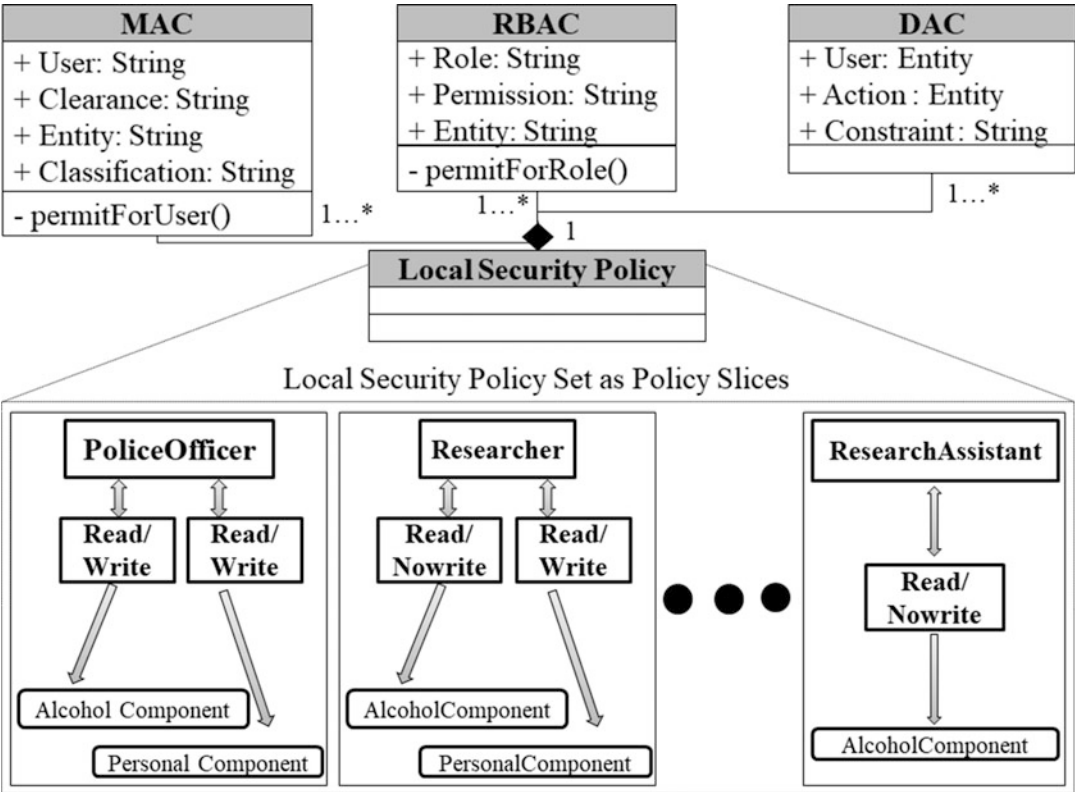
supports the big data application for different stakeholders. In support of these ideas, there is an associated three-step development process as shown in Fig. 19 that can be defined as follows:

1. Local systems have one or more security policies (or some specialized security policies) for the data they make available for use by the big data application. These security policies represent the data and permitted actions on the data.
2. A new big data application, like the FCR example in this entry, is developed and will utilize data from these local systems. In addition, the big data application will abide by the known local security policies.
3. The global security policy (GSP) for the big data application that is developed needs to integrate the known local policies that have been subscribed to and merge them appropriately with new security requirements.

We assume that each constituent system follows a publisher/subscriber model, facilitating the knowledge of what data comes from whom, and that each local security policy set has its own set of MAC classifications as an instance pair *<user, clearance>* for user representation and an instance pair *<entity, classification>* for entity classification, RBAC rules captured as instance triples *<role, permission, entity>*, DAC delegations as an instance triple *<user, action, constraint>*, and that the local system (e.g., State 1 in Fig. 1) may have multiple policies. We further assume that when given a set of local security policy sets, is it possible to match across



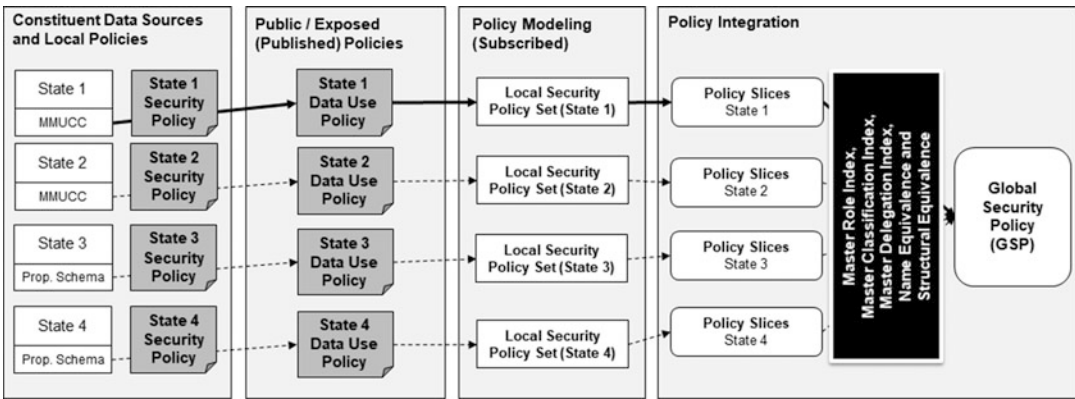
Access Control for XML Big Data Applications, Fig. 17 AD for User Van in the Law Enforcement Scenario



Access Control for XML Big Data Applications, Fig. 18 Example of a Local Security Policy Schema Set for the FCR Big Data Constituent System

them to determine which slices have some level of security equivalence. This is accomplished by performing name equivalence and structural equivalence operations on the security policy slices (Policy Integration box in Fig. 7).

As each constituent system establishes its own local security policy set, a Master Role Index tracks the definitions of new roles, ultimately expanding the shared role repository. Similarly, a Master Classification Index tracks the clearance



Access Control for XML Big Data Applications, Fig. 19 Processes to Integrate Local Security into Global Security for FCR

levels and classification of the data, and a Master Delegation Index tracks the delegation actions and constraints needed to be met.

As an example, State 1 or State 2 from Fig. 2 may only wish to expose some of local security policies, or slices of these policies, and may want to limit the security rules for each even further. For name equivalence, from Fig. 2, if State 1 has an RBAC policy slice with RoleID “1234” (e.g., Police Officer) with some role description A and role requirements B, and State 2 has an RBAC policy slice with Role “1234” with the same role description and role requirement values, then these are clearly equivalent by name. This is not enough information to conclude that they are also structurally equivalent. The reason is that the security rules are unique for each system; State 1 has its own set of security rules, and the ones for the Police Officer role might be different from the security rules for the Police Officer role in State 2. In order to see if they are the same, an algorithm for structural equivalence is executed. For structural equivalence, the RBAC rule hierarchies do not have to be identical, but the security rules that are constructed from each policy slice must be equivalent, namely:

$$\begin{aligned} & STATE_A.PolicySlice_{POLICEOFFICER}.SecurityRuleSet \\ & \equiv STATE_B.PolicySlice_{POLICEOFFICER}.SecurityRuleSe, \end{aligned}$$

The level of structural equivalence can vary, if for instance:

$$\begin{aligned} & STATE_A.PolicySlice_{POLICEOFFICER}.SecurityRuleSet \\ & \subset STATE_B.PolicySlice_{POLICEOFFICER}.SecurityRuleSet \end{aligned}$$

meaning that there are some rules in PoliceOfficer PSD for State 2 that are not present in State 1.

These equivalences at the structural level are crucial to provide security assurance in big data applications that might share data back to the constituent system. A GSP can be constructed for a big data application like the FCR example used in this entry by its subscription to the available local security policies of the constituent systems (states). Unlike integrating homogeneous policies (policies utilized on the same application, with roles, permissions, rules, etc., that target a limited, unchanged set of resources), our process aims to integrate heterogeneous policies (which contain equivalent and nonequivalent roles and permissions) that act on different resources utilized and provided by different local systems. To successfully create the GSP for a new big data application like FCR, the local security policy sets need to be integrated via similarity finding (name equivalence and structural equivalence) and other techniques that are able to analyze policy slices and their security rules across a set of published local security policies requested by the GSP.

Integration Process for Local Security Policy Sets

The integration process between multiple local security policy sets within FCR involves a name

equivalence utilizing the Master Role Index, the Master Classification Index, the Master Delegation Index, and structural equivalences (between security rules covering MAC, RBAC, and DAC). For example, in the case of RBAC, name equivalence verification between roles of heterogeneous local systems can be achieved by utilizing the Master Role Index, where two roles are name equivalent if and only if their respective name, description, and requirements are the same. Finding name equivalences results in two sets of roles. An equivalent set (*EqualRoles*), which satisfy the previous definition; and an unmatched set of roles (*DistinctRoles*), which do not match with any other roles. The policy slices that correspond to the roles in the set *DistinctRoles* can automatically become part of the new GSP, as there is no risk of conflict between other roles and their respective security rules. Note that two roles that are name equivalent may still have differing sets of security rules in their respective policy slices. As a result, structural equivalence must also be quantified.

Structural equivalence is a process that assumes name equivalence is met, and then proceeds to examine the actual security rules for each of the involved policy slices of local security policy sets. To verify structural equivalence, it is necessary to examine the security rules set of each involved policy slice. To do so, we note that there are four combinations that can exist as a result of the comparison between two security rule sets, regardless of whether they are MAC, RBAC, or DAC: equality, where the rule set of two systems are contained within one another; containment, where the rule set of one policy slice fully contains the rule set of another policy slice (proper containment, hence not equal); intersection, where the rule sets of two policy slices have at least one rule in common (intersection is not empty); and disjoint, where the rule set of two constituent systems with equivalent entities have no security rules in common (intersection is the empty set). These relations can be the result of either security rules overlapping with one another, or different entities being available as resources on one system, but not the rest (recall the security rule triple from section “[UML Extensions for Document Modeling](#)”).

The successful integration of these security rule sets, while maintaining a proper level of security that complies with the local security requirements of constituent systems, depends on the resulting relation between sets. Equality between sets is trivial; since no conflict can arise, both security rule sets can be consolidated and made part of the GSP. Handling rule sets that have a containment relation (proper subset) results in a subset of security rules being added to the GSP (all of the security rules of the contained smaller set), leaving the complement of the contained set in a state of conflict. Similarly, the intersection between two rule sets has two cases: when null, the rule sets are disjoint, and both rule sets are in a state of conflict and there are two options, either include them both or omit them both from the GSP; when non-null, only the intersection is included, requiring security rules that are not in the intersection to be excluded.

Resolving Conflicts of Integrated Security Rule Sets

Conflict resolution can be achieved by either utilizing a fully automated approach, which makes hard assumptions in order to provide the highest level of access control, or a human in-the-loop approach, where human intervention is necessary in order to select which security rules will become part of the GSP due to the intersection and disjoint cases. For an automated approach, we note that there exist different options capable of maintaining a proper sense of security:

- **Safe and lazy approach:** The GSP will utilize the most restrictive security rules, with respect of the entity involved, ensuring that no destructive operations (write) or a breach in access (reading) can be performed under MAC, RBAC, and DAC requirements.
- **Hierarchical approach:** A hierarchical order is given to the constituent local systems (and their local security policies) in order to decide which security rules takes precedence over other security rules.
- **Data Owner over Requester:** The GSP utilizes the security rules specified by the data

owner, disregarding other conflicting security rules.

While these automated approaches maintain a robust level of security, we note that they also present limitations. In the case of the safe and lazy approach, there could be automatic denials of access that could have otherwise been authorized in the local system via human intervention. The data owner approach presents the possibility of granting access to an otherwise secured request. In our example, State 1 can allow read permissions to Researchers in a localized manner (only when using the local system), but the big data application may want to limit the amount of reads to a subset of the information, or the operation itself throughout. The hierarchical approach offers the best alternative to conflict resolution, but the implementation depends highly on the order that the information published by local security policy sets would need to be ranked by some process (or human). While there are natural hierarchies of information and its granularity in domains such as health care, the same cannot be stated to other domains.

Clearly, the safest approach to conflict resolution is one where a human in-the-loop is employed in order to more precisely create the appropriate GSP for a big data application like FCR. This is not an unreasonable requirement, given our modeling and engineering approach to policy integration, since the human designer of the new application will be in the loop to examine the available policy slices for each local security policy set in order to decide which of them are relevant for the new application. This would require a final decision be made on all conflicting security rules. Potentially, if there is no name and structural equivalence between security rule sets, the designer will have to decide for each of the security rules that will make part of the system.

Creating the Global Security Policy

The person responsible of developing the big data application will use the capabilities described in

section “[Assumptions and Equivalence Finding](#)” and the potential integration of local security policy sets in section “[Integration Process for Local Security Policy Sets](#)” in order to arrive at a GSP that contains all of the relevant security rules, with respect to the roles, clearances/classifications, and delegations, packaged into a respective Global Security Policy (GSP). Using the approach, we decompose the local security policies by modeling them as security policy sets (higher level), policy slice (intermediate level), and security rule set (lower level); we build the GSP following a bottom-up approach. Since we assume that all of the security rule set conflicts have been resolved, whether by the use of a human-guided approach or an automated approach that satisfies the new system’s security requirements, the process of building the GSP involves three steps.

In the first step, Global Policy Slices are built utilizing the security rule sets with respect to roles, clearances/classifications, and delegations. This is done with the aid of the Master Role Index, the Master Classification Index, and the Master Delegation Index, ensuring that name equivalence and structural equivalence are correct. In the second step, a translation between the security rules represented in the policy slices (recall the triple $SR = \langle Role(r), Permission(p), Entities(ents) \rangle$ for RBAC, for example) and class members for the Global Policy Slices is performed. To achieve this, we group the security rules by Role and Entities, and propose the following equivalence:

- The combination of possible permissions (read, write) and Entities from the security rule will result in a class method with a stereotype that determines whether the method is permitted («allow») or denied («deny») for execution.

Once this step is completed, the third and final step is structuring the GSP artifact with respect to users and their clearances, roles, and delegations. An example on the integration with respect to roles is shown for the law enforcement big data

application example in Fig. 9. This involves several rules and steps:

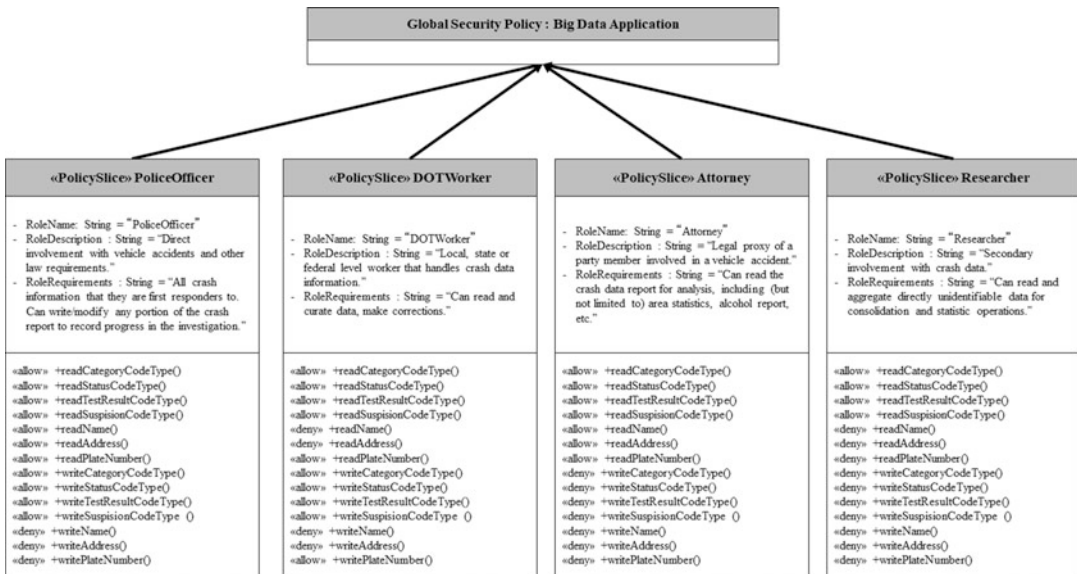
- Roles that will be part of the GSP are represented as subclasses of the GSP class.
- Each role class will have descriptive string members, *RoleDescription* and *RoleRequirements*.
- Each role class will have the methods, as given by the second step of the integration. This means that two methods per targeted entity of each role will be given, as dictated by the security rules.

As an example following RBAC security rules integration from constituent systems and the law enforcement scenario used throughout this entry, assume that multiple PoliceOfficer, DOTWorker, Attorney, and Researcher policy slices were integrated from State 1 and State 2 from Fig. 2. Roles are represented as subclasses of the parent GSP class of the big data application named as shown in Fig. 20. The subclasses named after the respective roles (PoliceOfficer, DOTWorker, Attorney, and Researcher) have two string members, which are instantiated to the Master Role Index values of the RoleDescription and RoleRequirements. In turn, these classes have one method per each security rule, and their permissions are given as

attached stereotypes. The DOTWorker and Attorney roles are structured similarly. We note that the structure of the GSP is not similar to the local security policy sets. The justification for this lies in the end use of each structure. In the case of the local security policy sets, it tries to model an already existing security policy in order to obtain the policy slices (and in turn, security rules) without the loss of structure or information. The GSP for the big data application is the result of integration that will be utilized as a blueprint for the generation of an actual enforcement policy. Due to this, we chose to structure the GSP in a role centric way, granting us the benefit of avoiding any complexity in the mapping process for the enforcement policy generation. Finally, the generation of enforcement security policy in the form of XACML code from the resulting GSP follows a similar process as the one utilized in our previous research (De la Rosa Algarín and Demurjian 2013), where XACML policies are generated from modeled Role Slice Diagrams.

Related Work

In this section, we describe and compare related research work done in three areas: big data security,



Access Control for XML Big Data Applications, Fig. 20 Representation of a Global Security Policy for FCR

policy and security modeling, and policy integration methods. With regard to big data security, security is often aimed at the data storage system, the back end of the system; while this information is important, we have focused on the front end of the users of the big data application. For example, security suggestions and approaches for big-data are often aimed at the data warehouses or storage systems (https://securosis.com/assets/library/reports/SecuringBigData_FINAL.pdf). Another common approach that relates big data with security is the exploitation of the former to improve the latter. Big data repositories can serve as knowledge bases to create more robust reactive and proactive security mechanisms (<http://codeascraft.com/2013/06/04/leveraging-big-data-to-create-more-secure-web-applications/>). As a matter of fact, the nature of the big data problem reduces the importance of security at the user level, focusing instead on performance and reliability of operations, as well as the development of applications with the sole purpose of efficient data exploration and discovery (<http://www.darkreading.com/applications/security-implications-of-big-data-strate/240151074>; <http://www.ibm.com/developerworks/library/bd-exploration/>). Krishnan (Krishnan 2014) presents two major challenges relating to access control and privacy policy specification in big data applications. The first challenge is the fine-grained access control to large-scale disparate data. As a potential solution, the authors specify that Attribute-Based Access Control (ABAC) could provide the necessary flexibility (<https://csrc.nist.gov/projects/abac/>). While our approach does not consider ABAC for constituent security requisites, the integration of MAC, RBAC, and DAC result in an equivalent alternative. Though the motivation for security in big data scenarios is very well defined, at the moment of this writing and with extensive literature research, we were not able to find any related work with big data security from the perspective of the application, e.g., the ability to define and control what the stakeholders of big data application are allowed to do at what times (<http://www.stanfordlawreview.org/online/privacy-paradox/big-data>; <http://unm2020.unm.edu/knowledgebase/technology-2020/14-are-you-ready-for-the-era-of-big-data-mckinsey-quarterly-11-10.pdf>). We believe that the work

presented in this entry is one of the earliest attempts at interpreting big data security from the perspective of the application and its stakeholders.

For security and policy modeling, one approach based on RBAC and UML, SecureUML (Lodderstedt et al. 2002), combines a graphical notation for RBAC with constraints, policies can be expressed using RBAC permissions, and complex security requirements can be done with a combination of authorization constraints. We take a similar graphical approach but concentrate on modeling and securing XML via the various diagrams presented in section “[Extensions for Document Security Policy Modeling and Integration](#).” A later effort by Basin et al. (2006) presents a model-driven security approach for designers to set security requirements along with the system models to automatically generate an access control infrastructure. The approach combines UML with a security modeling language defining a set of modeling transformations; the former produces infrastructures for JavaBeans, and the latter can generate secure infrastructures for web applications. Our work is done at a higher conceptual level. With respect to web service security, (Burmester et al. 2005) utilize the model-driven architecture paradigm to achieve security for e-government scenarios with inter-collaboration/communication. This is achieved by describing security requirements at a high-level (models), with relevant security artifacts being automatically generated for target architectures, removing the otherwise present learning curve in specifying security requirements by domain experts with no technical know-how. This is dramatically different from our work, which is focused on big data applications and its security assurance.

Policy integration approaches vary from similarity finding to specialized algebraic approaches. For example, (Mazzoleni et al. 2006, 2008) present a policy integration methodology to find similarity of policies on distinct levels (rule effects, targets, roles), and integrating over a set of defined rules, also considering policy decision conflicts. This is similar to our work, but differs in that our integration is done at the model level and is performed over homogeneous policies, while we have heterogeneous local security policy sets to be

integrated into a global security policy. Another approach (Rao et al. 2009) proposes an algebraic solution to the problem to integrate complex security policies at a granular detail with an algebra that consists of five unary operations (three binary and two unary). This algebra is utilized as part of a framework (to achieve the generation of an instance policy automatically. Like (Mazzoleni et al. 2008), this approach also focuses on the instance level, creating a dependency on the OASIS XACML specification and policy structure to achieve proper results (<https://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html>). The use of these methods would be impossible on security requirements defined in any other method (e.g., a database table with security rules, policies modeled with a different language, etc.).

Conclusion

In this entry, we have explored the attainment of mandatory, role-based, and discretionary access control for a big data application by presenting and discussing a framework for integrating local security policies of constituent systems to provide security assurance and facilitate the development of big data applications. To demonstrate the utility of the framework, we proposed a realistic scenario involving motor vehicle crash data collected from multiple states (see Fig. 2 again) that represents a Fusion of Big Data DaaS instances and Global Security Policy generation for a Federated crash repository (FCR). To achieve access control for big Data applications, the framework in this entry leverages UML as a document and policy-modeling tool, covering MAC, RBAC, and DAC concerns, and allows the creation of a global security policy to support the stakeholders of a big data application. Specifically, in section “[Use-Case Background](#)” we provided a brief background on the law enforcement and a MMUCC crash data use-case that served as the basis for Fig. 2, which presented a big data application use-case, the Federated Crash Repository including the underlying concepts, policy definition, and integration processes to link the local security policies (states) into a global context for the

crash data (country). Section “[UML Extensions for Document Modeling](#)” described the way that the Document Schema Class Diagram is used to represent XML using UML in support of FCR. Section “[Extensions for Document Security Policy Modeling and Integration](#)” provided an overview of the MAC, RBAC, and DAC UML extensions in our framework and exercised the framework with policy modeling and integration capabilities to capture local and global security policies in support of the security for the big data application FCR. Section “[Integrating Local Security Policies into a Global Security Policy](#)” detailed the process of integrating multiple local security policies into a global policy for the big data application FCR, including reconciling conflicts across local systems, roles, and permissions in order to attain a global policy that can be assured. To complete the work, section “[Related Work](#)” reviewed related research.

Bibliography

- Basin D, Doser J, Lodderstedt T (2006) Model driven security: from UML models to access control infrastructures. *ACM Trans Softw Eng Methodol* 15(1): 39–91
- Bell DE, La Padula LJ (1976) Secure computer system: Unified exposition and multics interpretation. MITRE CORP BEDFORD MA
- Burmester S, Giese H, Schäfer W (2005) Model-driven architecture for hard real-time systems: From platform independent models to code. *European Conference on Model Driven Architecture-Foundations and Applications*. Springer, Berlin, Heidelberg
- De la Rosa Algarín A (2014) An RBAC, LBAC and DAC Security Framework for Tree-Structured Documents. *Doctoral Dissertations*. 456
- De la Rosa Algarín A, Demurjian SA (2013) An Approach to Facilitate Security Assurance for Information Sharing and Exchange in Big-Data Applications. *Emerging Trends in ICT Security*. Morgan Kaufmann 2014:65–83
- De la Rosa Algarín A, et al. (2016) Securing XML with role-based access control: Case study in health care. *E-Health and Telemedicine: Concepts, Methodologies, Tools, and Applications*. IGI Global. 487–522
- De la Rosa Algarín A, et al. (2012) A security framework for XML schemas and documents for healthcare. 2012 *IEEE International Conference on Bioinformatics and Biomedicine Workshops*. IEEE
- De la Rosa Algarín A, et al. (2013a) Securing XML with role-based access control: Case study in health care. *E-*

- Health and Telemedicine: Concepts, Methodologies, Tools, and Applications. IGI Global 2016:487–522
- De la Rosa Algarín A, et al. (2013b) Generating XACML enforcement policies for role-based access control of XML documents. International Conference on Web Information Systems and Technologies. Springer, Berlin, Heidelberg, 2013
- De la Rosa Algarín A, Demurjian SA, Jackson E (2014) Access control for XML big data applications
- Ferraiolo DF, Sandhu R, Gavrila S, Kuhn DR, Chandramouli R (2001) Proposed NIST standard for role-based access control. *ACM Trans Inf Syst Secur (TISSEC)* 4:224–274
- Fowler M (2004) UML distilled: a brief guide to the standard object modeling language. Addison-Wesley Professional
- Guideline MMUCC (2012) Model Minimum Uniform Crash Criteria. DOT HS 811:631
- Krishnan R (2014) Access control and privacy policy challenges in big data. In: NSF Workshop on Big Data Security and Privacy, 2
- Lodderstedt T, Basin D, Doser J (2002) SecureUML: a UML-based modeling language for model-driven security. Springer, Berlin/New York, pp 426–441
- Mazzoleni, Pietro, et al. (2006) XACML policy integration algorithms: not to be confused with XACML policy combination algorithms!. Proceedings of the eleventh ACM symposium on Access control models and technologies
- Mazzoleni, Pietro, et al. (2008) XACML policy integration algorithms. *ACM Transactions on Information and System Security (TISSEC)* 11(1):1–29
- Pavlich-Mariscal JA, Demurjian SA, Michel LD (2010) A framework for security assurance of access control enforcement code. *Comput Secur* 29:770–784
- Rao P, et al. (2009) An algebra for fine-grained integration of XACML policies. Proceedings of the 14th ACM symposium on Access control models and technologies
- Sandhu RS, Samarati P (1994) Access control: principle and practice. *IEEE communications magazine* 32(9):40–48

A Theory of Approximate Arithmetic for Numerical Analysis: A Mathematical Foundation

A theory of approximate computing in numerical analysis

Tsau-Young Lin
Institute of Data Science, GrC Society, San Jose, CA, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Approximations, Metersticks, and Topologies
Approximate Arithmetic
Conclusions
References

Glossary

Numerical analysis A general term for this subject. Roughly speaking, it replaces manipulating real number by manipulating finite decimals.

Rounding In this chapter, the concept of regular rounding is replaced by the concept of measuring functions.

Measuring with a meterstick It is a geometric way of conducting rounding real numbers.

Vector Vector, not sequence, is used to name a linearly ordered set.

Definition of the Subject

Abstract: Numerical analysis, an ancient subject, lacks mathematical theory behind its practices. In this chapter, two basic concepts of

approximation are formalized. The first is the approximation defined by a mechanical procedure called measuring with a meterstick, such as when we use a meterstick to measure the diameter of a test tube. The mathematics behind this concept of approximation is called granular topology, which contains the usual topology as a sub-topology. The second concept of approximation is hidden in the approximate arithmetic. Since the classical approximate arithmetic is not well defined, a counter-example is provided and a new theory has been developed. This chapter summarizes what we have found. A mathematical version, with all the mathematical proofs, will be submitted to a journal of American Mathematical Society.

Introduction

Our interest in numerical analysis actually came from big data. Looking back, our words were quite intriguing, “In the age of big data, it is worthwhile to take a serious look at the fundamental idea of approximate computing” (Lin 2015); in big data any minute errors may become major issues. We were surprised to read “the effects of finite-precision arithmetic (a.k.a. round-off error) ... is not well understood at the most basic level, ... not a well-established mathematical model” (Scott 2021, Ch1, p. 1). The quoted words reinforce the need to examine the mathematical theory of approximate arithmetic. This chapter summarizes our findings and includes a model of approximate arithmetic, hopefully, which is a finite-precision arithmetic mentioned above.

In this introduction section, we will revisit two classical concepts of approximations, the two topologies related to measuring with a meterstick and approximate arithmetic. By a counter-example, we identify their inconsistent point and the errors of the classical model of approximations. First, two acronyms need to be introduced and defined:

Definition 1: Two acronyms

1. SRNC (Standard Rules/Practices in Numerical Computing) denotes the classical “theory” of numerical computing.
2. BTNC (Basic Theory of Numerical Computing) denotes a new theory of numerical computing that we are developing.

SRNC: Approximation for Numerical Computing

Before going into details, let us also define some abbreviations and notations:

Definition 2: Abbreviations and Notations

1. n , z , and x denote, respectively, the variables whose domains are non-negative integers, integers (zahlen), and the set, no topology, of real numbers $\mathcal{R}$; we assume these three variables will move through their respective domain automatically (as adopted from the practices of Einstein notation).
2. A half-open interval is a left-closed and right-open interval (Kelley 1955, p. 59).
3. n -dd is an abbreviation of a number with $\bar{n}$ decimal digits, and D_n is the set of all such n -dds.
 - Observe that an integer, say 3, is a 0-dd, which may also be represented by 3.0, a 1-dd, 3.00, a 2-dd, etc.
4. The term granule represents a neighborhood taken from a selected local base.

The primary goal of this subsection is to capture the approximation theory that has been used in SRNC. All SRNC-information is taken from the website (Measurement and Significant Figures). Some assertions (in the website prior to its latest updating) are quoted in the body of the reference (Some actual quotes from the website).

A meterstick has markings for centimeters (cm) and millimeters (mm). SRNC allows us to estimate one past the smallest marking, in this case millimeter, to the next decimal place in the measurement. Let us call such estimated markings fine-meters (theoretically, we assume that the markings for fine-meters are all precise). Using such a meterstick, we can measure, say, the

diameter $x_\infty = \frac{10}{3} = 3.33 \dots$ of a test tube to the nearest centimeter, millimeter, and fine-meter. Let us denote these measurements by $f_0 = 3$ cm, $f_1 = 3.3$ cm, and $f_2 = 3.33$ cm, respectively. Observe that if we vary x_∞ through $\mathcal{R}$, then these f_i for $i = 0, 1, 2$ mechanically define three measuring maps

$$m_i : \mathcal{R} \rightarrow D_i; x \mapsto m_i(x) \equiv f_i \text{ for } i = 0, 1, 2$$

SRNC (Measurement and Significant Figures) has not provided any formal concept of approximations. It simply uses measurements,

f_i for $i = 0, 1, 2$ as SRNC

$$\begin{aligned} &\text{-approximations of the real value } x_\infty \quad (1) \\ &= \frac{10}{3}. \end{aligned}$$

In section “[Approximations, Metersticks and Topologies](#)” we will generalize mathematically such an approximation “theory” into BTNC (namely, define a topological space).

For now, let us turn to a serious weakness of SRNC. (This counter-example was observed while we prepared a paper for IEEE Bigdata 2014; it is published in Lin (2015).)

SRNC: A Paradox? In Approximate Addition

In this subsection, we will show that SRNC-concept of approximate addition has an inconsistent point. Suppose we have two test tubes whose diameters are $x_\infty = \frac{10}{3}$ and $y_\infty = \frac{1}{3}$, respectively. Here are their measurements, $m_1(x_\infty) = 3.3$ and $m_2(y_\infty) = 0.33$, and implied mathematical relations:

$$\begin{aligned} x_\infty = \frac{10}{3} &\in m_1^{-1}(m_1(x_\infty)) = m_1^{-1}(3.3) \\ &= [3.25, 3.35 \equiv \sigma(m_1(x_\infty))]; \\ y_\infty = \frac{1}{3} &\in m_2^{-1}(m_2(y_\infty)) = m_2^{-1}(0.33) \\ &= [0.325, 0.335 \equiv \sigma(m_2(y_\infty))]. \end{aligned} \quad (2)$$

A1) SRNC: Approximate addition

Following SRNC (Measurement and Significant Figures), we compute this approximate addition in two steps:

1. The sum of approximations: $3.3 + 0.33 = 3.63$ and
2. $R_* : 3.63 \mapsto 3.6 = k$,

where R_* is a regular rounding operator.

SRNC concludes that (see some actual quotes from Measurement and Significant Figures [n.d.](#), Item 2):

A2) $k = 3.6$ is the SRNC-approximate sum.

Let us examine this SRNC-sum closely.

A3) Re-examining the SRNC

Suppose the two test tubes are glued perfectly together in the real world. Then the diameter of this new glued-tube is $x_\infty + y_\infty$. In addition, its measurement should be the (SRNC)-approximation of the exact sum. So the following equation, which is similar to Eq. (2), should be satisfied

$$\begin{aligned} x_\infty + y_\infty \in m_1^{-1}(k) &= [k - 0.05, k + 0.05] \\ &= [3.55, 3.65]. \end{aligned} \quad (3)$$

Unfortunately,

$$(x_\infty + y_\infty = 3.66 \dots) \notin m_1^{-1}(k) = [3.55, 3.65]. \quad (4)$$

In other words, Eq. (4) indicates that the SRNC-sum is not the approximation of the exact sum. It is clear that SRNC-approximation and SRNC-approximate additions do not form a coherent theory, So we need:

A4) A new theory

A new theory will be proposed in this chapter. Before we go into the theory, let us review and set up some notations.

Backgrounds: Lemmas, Notations, and Conventions

First let us set up some conventions and terminologies.

Definition 3: Conventions for using ϵ -neighborhoods, n -granules, and related terms

1. In college calculus, we call an open ball of radius ϵ in a k -dim Euclidean space ($\equiv \mathcal{R} \times \dots \times \mathcal{R}$) an ϵ neighborhood of the usual topology.
2. In this chapter, we will adopt a different set of such usual terms. An ϵ -neighborhood of the usual topology will mean an ϵ -open-cube, which is a Cartesian product of ϵ -open-intervals.
3. In addition, we will also use an ϵ -half-open cube as an ϵ -neighborhood, which is not an open neighborhood in the usual topology.
4. In a selected countable local base of half-open cubes, we will call a neighborhood an

$$n - \text{granule, } n - \text{cube, etc.} \quad (5)$$

if the length ϵ of half-open interval is $\epsilon = \frac{1}{10^n}$, where n is called the precision degree; observe that n automatically runs through non-negative integers (see Definition 2).

5. More generally, any neighborhoods in any selected local base will be called granules.

Note that all the terminology mentioned above *may* also be adopted for the granular topology $\mathcal{G}$. The following proposition taken, more or less, from Kelley (1955 Theorem 1, Ch3, p. 86) will be needed.

Proposition 1: A function $E(x)$ defined on a topological space X , which satisfies the first axiom of countability (Kelley 1955, p. 50) and has selected a countable local base at point x , is continuous at x if and only if for every granule N at $E(x)$, there is a granule M at x such that $E(M) \subseteq N$, where a granule is a general name of a neighborhood that is taken from a selected local base; see Definition 2.

Approximations, Metersticks, and Topologies

In this section, we will establish a mathematical theory of approximations for BTNC. This new theory will be built through some generalizations of SRNC-practices (Measurement and Significant Figures) mentioned above.

1. Our first generalization is to replace “the meterstick that has markings on centimeters” with a real line, called an n -line which has markings on n -dds, where n , an Einstein notation, varies through every non-negative integer; see Definition 2. Observe that we will assume, as in college calculus, the real numbers are points on the real line and vice versa.
2. The second generalization is to replace the mechanical procedure, measuring with a meterstick, with a geometric procedure, measuring with an n -line. Here are the details. Based on the illustration of Fig. 1.8.1 in (Measurement and Significant Figures), we will do the following translation. We locate (a) x_∞ on an n -line (instead of on a meterstick) and find (b) the nearest n -dd on this n -line (instead of the nearest centimeter on a meterstick).

Since “measuring with a meterstick” is not a mathematical procedure, our formalization and generalization can not be a piece of formal mathematics, even though its “final outputs,” measuring functions are. Our formalization, though is not a formal mathematical reasoning, hopefully is a so-called compelling arguments; see, for example Hopcroft and Ullman (1979, p. 147). Anyway in BTNC, we will assume that we have the following mathematical measuring functions m_n :

$$m_n : R \rightarrow D_n; \quad x \mapsto m_n(x). \quad (6)$$

Since n automatically varies through every non-negative integer (Definition 2), we also have

$$m_* : R \rightarrow \{D_n \mid n\}; \quad x \mapsto \{m_n(x) \mid n\}. \quad (7)$$

We will call $m_*(x) = \{m_n(x) \mid n\}$ in Eq. (7) a vector of (formalized) measurements of x and its

n -th component $m_n(x)$ n -th measurement. (We choose “vector” to name an ordered set since “sequence” may imply that convergency is based on the usual topology.)

At this point, it is important to point out a hidden convention. Let us explain it by an example.

Example 1: A Convention

Suppose we are given a real number $x_\infty = \frac{3}{10}$. Then x_∞ has two nearest 2-dds, 3.25 and 3.35 whether on an n -line or on a meterstick. In such a situation, SRNC has adopted a convention that chooses the smaller one 3.25 to be the *unique* nearest 2-dd, namely, $m_2(x_\infty) = 3.25$.

- We will adopt such a convention throughout the whole chapter.

These measuring functions will be our base of BTNC-approximation theory.

Measuring Functions and Topologies

With these preliminaries, we are ready to present a formal approximation theory that is based on the vector of measuring functions defined in Eq. (6) and Eq. (7). Observe that by the theory explained in ((Kelley 1955, p. 96, or Stanat and McAllister 1977, p. 211 and p. 185), the vector of measuring functions induces a vector $P = \{P_n \mid n\}$ of partitions (see Definition 2 about n); let its union be $P_\infty = \cup P$. (We choose “vector” to name an ordered set since “sequence” may imply that convergency is based on the usual topology.)

- In an area of AI, called Rough Set theory (Pawlak 1982), the “standard” universes of discourse are topological spaces (Lin and Liu 1994) generated by some partitions that are often called classifications and/or clusterings in AI. So $P = \{P_n \mid n\}$ is a vector of rough set topologies and we are interested in the “limit” of rough set topologies; see $\mathcal{P}$ below. For our goal, we need to re-organize P into a vector of refining partitions (for the term “refinement,” see Kelley 1955, p. 128, or Stanat and McAllister 1977, Definition 3.7.6, p. 185).

$$G = \{G_n = P_0 \cap \dots \cap P_n \mid n\} \text{ and} \quad (8)$$

$$G_\infty = \bigcup G \quad (9)$$

Now, we will re-state Theorem 12 in (Kelley 1955, p. 47) in terms of P_∞ and G_∞ as our first theorem.

Theorem 1: Granular Topology

Let S_∞ be P_∞ or G_∞ . “If S_∞ is any non-void family of sets the family B^S of all finite intersections of members of S_∞ is the base for a topology S for the set $R = \bigcup \{s \mid s \in S_\infty\}$,” the set of all real numbers. Moreover, the two topologies $G = P$ and $B^G = G_\infty$. (This quote is a very precise quote, except for the symbols.)

Proof: The main assertion follows from original “Theorem 12.” Here are the proofs of two additional assertions. Note that a finite intersection of members of G_∞ is a finite intersection of members P_∞ , since every member of G_∞ is a finite intersection of P_∞ . Thus, we have $B^G \subseteq B^P$ that implies $\mathcal{G} \subseteq \mathcal{P}$. Conversely every member of P_∞ can be expressed as a union of members of G_∞ , and hence B^G can generate B^P that implies $\mathcal{P} \subseteq \mathcal{G}$. Hence the two topologies are equal $\mathcal{G} = \mathcal{P}$. Let us turn to the last assertion. Note that a member of B^G , which is a finite intersection of members of G_∞ , is a member of $P_0 \cap \dots \cap P_n = G_n$ for some n and hence is a member of G_∞ (observed that G is a sequence of refining partitions); in other words, $B^G = G_\infty$. QED.

Theorem 2: The family,

$$B^G(x) = \{b_n^G(x) \mid x \in b_n^G(x) \in G_n\} \quad (10)$$

is a local base $B^G(x)$ at x , where $b_n^G(x)$ is the unique equivalence class in G_n that contains x .

Theorem 3: The measurement vector $m_(x) = \{m_n(x) \mid n\}$ converges to x relative to the usual topology $\mathcal{U}$. Moreover, the following map*

$$m_* : \{m_*(x) \mid x \in \mathcal{R}\} \rightarrow \mathcal{R} \quad (11)$$

is one-to-one and onto.

Theorem 4: The identity map $i : \mathcal{R}_G \rightarrow \mathcal{R}_U$ is continuous (by Proposition 1) and $i^{-1}(\mathcal{U})$, a homeomorphic copy of the usual topology, is a sub-topology of $\mathcal{G}$.

Approximate Arithmetic

So far, we have established an approximation theory for BTNC. Now we turn to the scientific computation part of numerical analysis, which is “the wide spectrum of activities having to do with the approximate solution of scientific problems expressed through mathematical models” (Linz 1988). In other words, the approximation theory for scientific computing is real numbers based (with the usual topology). In order to coordinate the approximation theory with the structures of real numbers, we will focus on measurement vectors m_*^U , which is a substructure of m_* . To that end, we will require that m_*^U be a U -homeomorphism and introduce $\oplus$ and $\otimes$ as follows.

- The whole section, except the section “[Isomorphisms of Ordered Fields](#)”, all concepts and theorems are valid for granular topology too. In other words, except the inequalities, all theorems are also valid for G .

Definition 4:

$$[m_*^U(x) \oplus m_*^U(y)]_p \equiv [m_*^U(x+y)]_p \text{ and} \quad (12)$$

$$[m_*^U(x) \otimes m_*^U(y)]_p \equiv [m_*^U(x \times y)]_p \quad \forall p. \quad (13)$$

Thus two semi-groups $(\mathcal{D}, \oplus)$ and $(\mathcal{D}, \otimes)$ are formed.

Semi-group $(\mathcal{D}, \otimes)$

Abstractly speaking, these two semi-groups are the “same.” So, we will explore the harder one closely and state the result on the easier one without going through tedious proof.

Before we go into the details, let us recall a common “abusing” of notations. Suppose we have been given the function of multiplication, $E(x, y) = x \times y$. We often extend the E to the following situations.

Definition 5: Abused notations

$$E(A, B) = \{a \times b \mid a \in A, \text{ and } b \in B\}$$

where A and B be two subsets of the reals (Kelley 1955, p. 11).

Lemma 1: Semi-group $(\mathcal{D}, \otimes)$

Let E be the continuous function that defines the multiplication “ $\times$ ” on $\mathcal{R}_U$. Let $E(x_\infty, y_\infty) = x_\infty \times y_\infty$. Then by Proposition 1, for every integer $p \geq 0$ (that implies $\exists p$ -granule $b_p(x_\infty \times y_\infty)$ in the local base), there is a minimal $n(p) \geq p$ (that implies $\exists$ two maximal $n(p)$ -granules $b_{n(p)}(x_\infty)$ and $b_{n(p)}(y_\infty)$ in their respective local bases) such that (The E in $E(b_{n(p)}(x_\infty), b_{n(p)}(y_\infty))$ is an abused notation.)

$$\begin{aligned} E(b_{n(p)}(x_\infty), b_{n(p)}(y_\infty)) \\ = b_{n(p)}(x_\infty) \times b_{n(p)}(y_\infty) \\ \subseteq b_p(E(x_\infty, y_\infty)) (\text{and } \Rightarrow) \end{aligned} \quad (14)$$

$$\begin{aligned} m_p^U(m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty)) \\ = [m_*^U(x_\infty) \otimes m_*^U(y_\infty)]_p \\ = [m_*^U(x_\infty \times y_\infty)]_p \equiv m_p^U(x_\infty \times y_\infty) \end{aligned} \quad (15)$$

where m_n^U is a measuring function, and the two granules are $b_p(t) = (m_p^U)^{-1}(m_p^U(t))$, and $b_{n(p)}(t) = (m_{n(p)}^U)^{-1}(m_{n(p)}^U(t))$.

- This Lemma can be more general; E can be a continuous function of m -variables.

Proof of Lemma 1:

Since E is continuous at $(x_\infty \times y_\infty)$, by Proposition 1, for any given p (That implies $\exists p$ -granule $b_p(x_\infty \times y_\infty)$ in the local base), there is a minimal $n(p)$ (That implies $\exists$ two maximal $n(p)$ -granules $b_{n(p)}(x_\infty)$ and $b_{n(p)}(y_\infty)$ in their respective local bases) such that

$$b_{n(p)}(x_\infty) \times b_{n(p)}(y_\infty) \subseteq b_p(x_\infty \times y_\infty).$$

So we have established Eq. (14). Since $m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty)$ is located in $b_{n(p)}(x_\infty) \times b_{n(p)}(y_\infty)$, by last equation, we have

$$\begin{aligned} m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty) \in b_{n(p)}(x_\infty) \times b_{n(p)}(y_\infty) \\ \subseteq b_p(x_\infty \times y_\infty) (\Rightarrow). \end{aligned} \quad (16)$$

By applying m_p^U to Eq. (16) and noting that $b_p(x_\infty \times y_\infty) \equiv (m_p^U)^{-1}(m_p^U(x_\infty \times y_\infty))$, so we have

$$\begin{aligned} m_p^U(m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty)) \\ = m_p^U(b_p(x_\infty \times y_\infty)) \\ = m_p^U(m_p^U)^{-1}(m_p^U(x_\infty \times y_\infty)). \end{aligned} \quad (17)$$

The first equality of Eq. (18) is derived from the last term of Eq. (17).

$$\begin{aligned} m_p^U(m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty)) \\ = m_p^U(x_\infty \times y_\infty) \equiv [m_*^U(x_\infty \times y_\infty)]_p. \end{aligned} \quad (18)$$

Using Eq. (13), we transform the last equation to

$$\begin{aligned} m_p^U(x_\infty \times y_\infty) \equiv [m_*^U(x_\infty \times y_\infty)]_p = \\ [m_*^U(x_\infty) \otimes m_*^U(y_\infty)]_p \\ = m_p^U(m_{n(p)}^U(x_\infty) \times m_{n(p)}^U(y_\infty)), \end{aligned}$$

which is Eq. (15) in reverse order. QED.

Corollary 1:

$$\begin{aligned} m_p^U(x \times y) = m_p^U(m_{n(p)}^U(x) \times m_{n(p)}^U(y)) \\ \forall p \geq 0. \end{aligned} \quad (19)$$

- This is the key theorem for computing the “approximate multiplication.”

Semi-group $(\mathcal{D}, \oplus)$

Observe that the “same proof” works for Lemma 2 and corollary 2:

Lemma 2: Semi-group $(\mathcal{D}, \oplus)$

Let $E(x, y)$ be the continuous function of the addition on the reals $\mathcal{R}_G$, and let $E(x_\infty, y_\infty) = x_\infty + y_\infty$. Then, for every integer $p \geq 0$, there is a minimal $n(p) \geq p$ such that

$$\begin{aligned} E(b_{n(p)}(x), b_{n(p)}(y)) = b_{n(p)}(x) + b_{n(p)}(y) \\ \subseteq b_p(E(x, y)) (\text{and } \Rightarrow) \end{aligned} \quad (20)$$

$$\begin{aligned}
m_p^U \left(m_{n(p)}^U(x) + m_{n(p)}^U(y) \right) \\
&= [m_*^U(x) \oplus m_*^U(y)]_p \equiv [m_*^U(x+y)]_p \\
&= m_p^U(x+y).
\end{aligned} \tag{21}$$

Corollary 2:

$$\begin{aligned}
[m_*^U(x) \oplus m_*^U(y)]_p &= m_p^U(x+y) \\
&= m_p^U \left(m_{n(p)}^U(x) + m_{n(p)}^U(y) \right) \forall p \geq 0
\end{aligned}$$

Topological Fields

In the last section, we have developed two semi-groups. Based on that, we will extend the measurement vector m_*^U into two isomorphisms of two topological fields.

To that end, let us introduce:

Definition 6: Bi-operator algebras

$(\mathcal{X}, \oplus, \otimes)$ is called bi-operator algebras, if $(\mathcal{X}, \oplus)$ and $(\mathcal{X}, \otimes)$ are semi-groups.

So $(\mathcal{D}, \oplus, \otimes)$ are bi-operator algebras. By combining these together, we have an isomorphism of bi-operator algebras.

$$m_*^U : (\mathcal{R}, +, \times) \rightarrow (\mathcal{D}, \oplus, \otimes) \tag{22}$$

Again, let us have another easy isomorphism of fields:

$$m_*^U : (\mathcal{R}, +, \times) \rightarrow (\mathcal{D}, \oplus, \otimes), \tag{23}$$

In his book, Jacobson (1985, p. 49) states that a ring is a bi-operator algebra plus one additional axiom, called the distributive law. Since (m_*^U) is one-to-one and onto map, this new map will import the distributive law from $(\mathcal{R}, +, \times)$ into $(\mathcal{D}, \oplus, \otimes)$. So we have proved this isomorphism.

An important next step is to extend Eq. (23) to two isomorphisms of topological fields.

Theorem 6:

$$m_*^U : (\mathcal{R}, +, \times, \mathcal{U}) \rightarrow (\mathcal{D}, \oplus, \otimes, \mathcal{T}_1) \tag{24}$$

$$m_* : (\mathcal{R}, +, \times, G) \rightarrow (\mathcal{D}, \oplus, \otimes, \mathcal{T}_2) \tag{25}$$

are isomorphisms of topological fields.

Proof: (1) This map m_*^U defined in Eq. (23) will import the topology $\mathcal{U}$ on $(\mathcal{R}, +, \times)$ to $(\mathcal{D}, \oplus, \otimes)$. After the topology is ported, we have this theorem.

(2) A similar proof for m_* . QED.

Isomorphisms of Ordered Fields

Theorem 7:

$$m_*^U : (\mathcal{R}, +, \times, >) \cong (\mathcal{D}, \oplus, \otimes, >) \tag{26}$$

is an isomorphism of ordered field.

Proof: In Theorem 6, we have shown that the map m_*^U is a homeomorphism and an isomorphism of fields. Observe that since the usual topology is an order topology (Kelley 1955, p. 58, J(b)). Hence the homeomorphism preserves the order. So we have this Theorem. QED.

Theorem 8:

$$\begin{aligned}
[m_*^U(x+y)]_p &= m_p^U(x+y) \\
&= m_p^U \left(m_{n(p)}^U(x) + m_{n(p)}^U(y) \right).
\end{aligned} \tag{27}$$

$$\begin{aligned}
[m_*^U(x \times y)]_p &= m_p^U(x \times y) \\
&= m_p^U \left(m_{n(p)}^U(x) \times m_{n(p)}^U(y) \right) \forall p \geq 0.
\end{aligned} \tag{28}$$

1. Theorem 8 is about (1) approximate $E(x, y) = x + y$ and (2) approximate $E(x, y) = x \times y$.
2. The theorems in this section may not be valid for granular topology $\mathcal{G}$.
3. Here is an interesting question: Can Theorem 8 be generalized to (3) approximate any continuous functions of n -variables (n is in Einstein notation)?

Conclusions

In this chapter, we have (1) examined the SRNC-theory, (2) identified its errors, and (3) provided a

basic theory of numerical computing (BTNC). Here are some of them. (1) We have introduced granular topology to replace the mechanical procedure called measuring with a meterstick. (2) We have found some interactions between granular topology and usual topology. Here is an interesting view that we will address it in the full mathematical view of this paper in an AMS journal: By regarding a local base as a variable interval, we can formulate a theory of variable interval computing along the line of Kearfott (n.d.); see Lin (2019). By Theorem 7, we may conjecture that we should be able to transform a Smale computable function (Blum et al. 1989) into a vector of convergent Turing computable function (Hopcroft and Ullman 1979) on D .

References

- Blum L, Shub M, Smale S (1989) On a theory of computation and complexity over the real numbers: NP-completeness, recursive-functions, and universal machines. *Bull Amer Math Soc* 21(1989):1–46
- Hopcroft J, Ullman J (1979) Introduction to automata theory, languages, and computation. Addison-Wesley Publishing Company
- Jacobson N (1985) Abstract algebra. Dover Publication
- Kearfott RB. Interval computations: introduction, uses, and resources. <https://interval.louisiana.edu/preprints/survey.pdf>
- Kelley J (1955) General topology. Van Nostrand. Ch 3, Theorem 1, Item e
- Lin TY (2015) A paradox in rounding errors approximate computing for big data. *SMC* 2015:2567–2573
- Lin TY (2019) Approximate computing in numerical analysis variable interval computing – extended abstract. *SMC* 2019:2013–2018
- Lin TY, Liu Q (1994) Rough approximate operators-axiomatic. In: Ziarko W (ed) *Rough set theory, rough sets, fuzzy sets and knowledge discovery*. Springer-Verlag, pp 256–260. (Co-author: Qing Liu)
- Linz P (1988) A critique of numerical analysis. *Bulletin (New Series) of AMS* 19(2)
- Pawlak Z (1982) Rough sets. *Int J Parallel Prog* 11(5): 341–356
- Scott LR (2021) *Numerical analysis*. Princeton University Press, 2011
- Measurement and Significant Figures (n.d.). [https://chem.libretexts.org/Bookshelves/Introductory_Chemistry/Map%3A_Fundamentals_ofGeneral_Organic_and_Biological_Chemistry_\(McMurry_et_al.\)/1%3A_Matter_and_Measurements/1.09%3A_Measurement_and_Significant_Figures](https://chem.libretexts.org/Bookshelves/Introductory_Chemistry/Map%3A_Fundamentals_ofGeneral_Organic_and_Biological_Chemistry_(McMurry_et_al.)/1%3A_Matter_and_Measurements/1.09%3A_Measurement_and_Significant_Figures). Some actual quotes from the website:
- 1) “In a measurement, all of the certain digits (those read directly from the scale) and one uncertain digit (one number made by estimating between marks on the scale) are significant”
 - 2) “Remember that calculators do not understand significant figures. You are the one who must apply the rules of significant figures to a result from your calculator”
 - 3) “Assigning significant figures and using the rules to round the result of computations gives us a systematic way of indicating the degree of trustworthiness of the result”
 - 4) “The foundation of significant figures is preserving the degree of uncertainty of the instruments”
- Stanat D, McAllister D (1977) *Discrete mathematics in computer science*. Prentice-Hall

Interval Computing

Ralph Baker Kearfott
Department of Mathematics, University of
Louisiana, Lafayette, LA, USA

Article Outline

Definition of Interval Computing
Introduction
Key Information
Future Directions
References

Glossary

Branch and bound algorithms algorithms that provide some kind of exhaustive search for an optimum value of a quantity, by subdividing a domain, storing the parts resulting from the subdivision, bounding the quantity over the parts, and, based on the bounds over a part, discarding the part, recursively subdividing the part, or registering the search as being completed over that part.

Constraint propagation given intervals or sets in which values must lie, the use of equations, inequalities, logical expressions, and other information involving those values to reduce the width of the intervals or sizes of the sets.

Directed rounding in a system for computer arithmetic, rounding an exact number in a particular direction (up or down) to approximate the number from a set of numbers with a particular finite number of digits.

Global optimization the search for the best possible value over the *entire* feasible set (i.e., the entire set of possibilities), as opposed to finding values that are merely better than similar or nearby values.

Interval arithmetic arithmetic defined on intervals of real numbers, such that the result of an

operation contains the range of real values over the interval arguments.

Mathematically rigorous capable of providing a proof accepted by mathematicians as complete, self-standing, and without further qualification, as opposed to providing mere empirical evidence.

Round off error the error in a computation resulting from approximating a real number by an expansion (decimal, binary, or other base) with a finite number of digits.

Subdistributivity the property in a system of arithmetic containing an addition and multiplication that, for objects a , b , and c upon which the addition and multiplication is defined, $a(b + c) \subseteq ab + ac$, but $a(b + c)$ is not necessarily equal to $ab + ac$.

Wrapping effect the explosive overestimation in the interval vector enclosure of the dependent variable vector in a dynamical system due to the fact that the image of an interval vector after a time step is not an interval vector, but is enclosed in one.

Definition of Interval Computing

Here, we take interval computing to mean the use of arithmetic based on intervals or on directed rounding, to handle uncertainties in measurements and encompass round-off error in floating-point computations, in applications such as encompassing the range of possible system behaviors, in automatic theorem proving, in producing foolproof systems in the presence of uncertainty, etc.

Introduction

We give motivation for the subject, a historical outline, and the basic idea here, while we give details, applications, and a description of resources in subsequent sections.

Motivation

Several motivations underlie the development and use of interval arithmetic. These include

- Determining all possible behaviors of a system with known bounds on uncertainties in design parameters
- Guaranteeing bounds on worst case behavior of systems that cannot fail, given bounds on the parameters in the system
- Automating round off error analysis in floating-point arithmetic
- Computing quick but mathematically rigorous bounds on the range of quantities, for use within larger computations

The mathematical concept underlying traditional interval arithmetic is simple and has been discovered multiple times over at least a century. Namely, given two intervals $x = [\underline{x}, \bar{x}]$ and $y = [\underline{y}, \bar{y}]$ define each elementary operation $\odot$, $\odot \in \{+, -, \times, \div\}$, so that the result of $x \odot y$ is the set of all possible results $x \odot y$ for $x \in x$ and $y \in y$, that is,

$$x \odot y = \{x \odot y \mid x \in x \text{ and } y \in y\}. \quad (1)$$

Early History

Equation 1 leads to operations that are nearly as simple to implement as complex arithmetic, although there are complications that we explain in subsequent sections. Perhaps because this is a basic and natural idea, classical interval arithmetic has been developed independently many times. Siegfried Rump mentions evidence in newsletters for high-school mathematics teachers in nineteenth-century Germany that a type of interval arithmetic was taught, as common knowledge. An example of this is Schnellinger (1879), where Schnellinger describes in great detail over- and under-estimates and determination of the number of significant digits when using truncated decimal arithmetic; Schnellinger attributes the importance of such arithmetic to the introduction of the metric system. Rump mentions French work from 1887, 1879, and 1854 in which formulas for the

elementary operations and rigorous error bounds are given; an early one is (Burat 1885, Complément).

Rump even claims Gauss provided explicit computation of error bounds, including rounding errors, in an 1809 work. Presumably this is (Gauss 1809, Book 1, Section 1), where Gauss analyzes the error in approximating transcendental functions, for use in accurately determining the orbits of planets. We will leave it to a reader with a better knowledge of classical Latin to verify this.

Modern History

In the twentieth century, an early paper is Young (1931). The operational definitions are essentially those used today, although the motivation was that of theoretical mathematical analysis; that is, Young viewed it as an arithmetic of limits, where $\liminf_{x \rightarrow x_0} f(x) \neq \limsup_{x \rightarrow x_0} f(x)$.

Several papers appeared almost simultaneously but, it is claimed, independently, with the advent of modern digital computers. In Dwyer (1951), Dwyer introduces interval computations as an integral part of round-off error analysis. In Warmus and Steinhaus (1956), Warmus develops interval arithmetic as a means to provide a sound theoretical backing to numerical computation. In Sunaga (1958), Sunaga presents interval arithmetic in its essentially currently recognizable form, emphasizing mathematical theory but motivated by accounting for uncertainty and error in measurement and computation.

Interval mathematics in its current form is largely attributed to Ramon Moore, originating from work done at Lockheed Missiles and Space Company in possible collaboration with Eldon Hansen. He first described his work in Moore (1959) (possibly independently from Sunaga), and later, fully, in his dissertation Moore (1962). Moore continued to explain and advocate the use of interval arithmetic and associated algorithms, largely through the two editions of his monograph Moore (1966, 1979), and our more recent update Moore et al. (2009).

Notable centers of study and development arose. Some, but not all, examples of early or

significant literature in Europe include Nickel (1966); Apostolatos and Kulisch (1967a,b); Henrici (1971); Nickel (1973); Alefeld and Herzberger (1974); Kulisch and Miranker (1981); Rump (1981); Rohn (1981). Karl Nickel oversaw the *Freiburger Intervallberichte*, an extensive institutional series containing many seminal ideas from recognized researchers in the subject. Somewhat later, Russian researchers started the journal *Interval Computations*, later changing its name to *Reliable Computing* and published by Kluwer, then acquired by Springer; it continues today as a noncommercial open-access online journal.

Biennial SCAN (Scientific Computing, Computer Arithmetic, and Validated Numerics) meetings featuring interval computations have been held for several decades, and smaller regional conferences are held in intervening years. SCAN proceedings typically occur in special journal volumes and lecture note series, such as Hansen (1969); Nickel (1975); Rao and Matula (1975); Alefeld and Grigorieff (1980); Nickel (1980); Kulisch and Miranker (1983); Miranker and Toupin (1985); Nickel (1986); Kulisch and Stetter (1988); Kaucher et al. (1991); Atanassova and Herzberger (1992); Adams and Kulisch (1993); Corliss and Kearfott (1993); Herzberger (1994); Knowles and McAllister (1995); Alefeld and Herzberger (1996); Alefeld et al. (1996); Kearfott and Kreinovich (1996); Csendes (1999); Krämer and von Gudenberg (2001); Alt et al. (2004); Luther and Otten (2006); Ceberio and Kreinovich (2008); Shary and Corliss (2012); Nehmeier et al. (2016); Kreinovich and Tucker (2012).

Interval computations have also been featured in other proceedings, such as those related to more general uncertainty analysis, computer arithmetic, global optimization, constraint programming, computer graphics, automatic differentiation, fuzzy technology, computer algebra, and various applied application areas. Additional research has appeared in the University of Wisconsin Mathematics Research Center (MRC) technical report series.

Key Information

Basic Definitions and Properties

The logical operational definitions (1) lead naturally to these operational definitions.

$$\left. \begin{aligned} x + y &= [\underline{x} + \underline{y}, \bar{x} + \bar{y}], \\ x - y &= [\underline{x} - \bar{y}, \bar{x} - \underline{y}], \\ x \times y &= \left[\min \{ \underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y} \}, \max \{ \underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y} \} \right] \\ \frac{1}{x} &= \left[\frac{1}{\bar{x}}, \frac{1}{\underline{x}} \right] \text{ if } \underline{x} > 0 \text{ or } \bar{x} < 0 \\ x \div y &= x \times \frac{1}{y}. \end{aligned} \right\} \quad (2)$$

Equations (2) lead to efficient and automatic computations with interval arithmetic, either in hardware or with operator overloading. Depending on the efficiency of multiplication and comparison operations, Eqs. (2) directly lead to efficient computations on many platforms, while, on some, implementation of multiplication with a chain of nine cases (see, e.g., Moore et al. 2009, Table 2.1, p. 13) and separate consideration of the division operation are most efficient.

Efficient machine implementations of operations (2) are done with floating-point arithmetic, but, as such, can only approximately satisfy the logical condition (1). Some practitioners in some applications, such as fuzzy technology, deem such approximate operations to be sufficient. However, to use interval arithmetic in automatic theorem proving, to prove the correctness of computer code, or in applications where all possible outcomes with uncertain parameters must be known with mathematical rigor, more is needed.

Based on the logical definition (1), various kinds of extended interval arithmetic has been developed. If $\odot^\approx$ is a floating-point implementation of interval arithmetic, it is sufficient that

$$x \odot^\approx y \subseteq \{x \odot y \mid x \in x \text{ and } y \in y\}. \quad (3)$$

With *directed rounding*, that is, rounding the computed lower bound of an operation obtained with (2) down and the computed upper bound up, the condition in (3) is efficiently achieved. The IEEE 754-2019 floating-point arithmetic standard IEEE-754 (2019), as well as all earlier versions of this standard, specifies directed rounding, implemented in the most commonly available desktop computer chips and programming languages.

Common functions (square root, exponential and trigonometric functions, etc.) are treated as elementary operations, with values defined according to (1). Machine interval evaluation of such functions is done by combining usual approximation methods with directed rounding or known computational error to compute lower and upper bounds.

When a denominator in interval division contains zero, the logical condition (1) leads to operational definitions, defining empty, semi-infinite, and infinite intervals, in *extended interval arithmetic*. Such extended arithmetic's contain subtle differences depending on whether infinity is considered part of the number system. Similarly, interval arithmetic's and extended interval arithmetic's differ depending on how evaluation over intervals only partially within the domain is handled; for example, in some applications it is reasonable to define $\sqrt{[-1, 4]}$ to be $[0, 2]$, whereas the negative values need to be noted in other applications. The IEEE 1788-2015 standard for interval arithmetic IEEE (2015) and the subset standard IEEE (2017) specify *decorations* associated with the interval datum that track various conditions that occur during a chain of interval operations.

Although the logical condition (1) is exactly satisfied when the operational formulas (2) and are satisfied to within a unit round-off error for optimally implemented floating-point-based operations $\odot^\approx$, this is not so for expressions composed of many operations. In particular, for intervals x , y , and z , only *subdistributivity* holds:

Proposition 1 *If x , y , and z are intervals, then*

$$x(y + z) \subseteq xy + xz. \quad (4)$$

Furthermore, $x(y + z)$ and $xy + xz$ are equal, that is, they are the same sets only under certain conditions.

Example 1 If $x = [-1, 2]$, $y = [1, 3]$, and $z = [-3, 1]$, we have $x(y + z) = [-4, 8]$, while $xy + xz = [-9, 9]$. However, if we define $f(t, u, y, z) = ty + uz$, then $[-9, 9]$ is the exact range of f over the domain defined by $t \in [-1, 2]$, $u \in [-1, 2]$, $y \in [1, 3]$, $z \in [-3, 1]$.

In the expression $xy + xz$, the interval arithmetic implicitly assumes the values x in the first term vary independently from the values x in the second term, an assumption that is not always met in practice. The resulting overestimation is not catastrophic in some applications but can be in others. Naive application of interval arithmetic to complicated expressions or to large computer codes that evaluate quantities is likely to lead to the useless bounds $(-\infty, \infty)$, which, although correctly bounding the range, are unhelpful. Some ways of getting around this are:

- Preprocessing expressions and codes to reduce the number of occurrences of variables
- Using point floating-point operations to obtain good guesses on results, applying convergent interval iterative processes strategically to verify rigorous bounds on true results afterwards
- Dispensing with working with entire intervals, redesigning the computation to use directed rounding's directly to separately compute lower and upper bounds on the result

Such techniques are unnecessary for various simple but important computations where bounds on the range are needed.

There are additional pitfalls if one wishes to bound the range of possibilities of a numerical algorithm. For instance, the *wrapping effect*, first described in Ramon Moore's dissertation, is well known among developers of software for interval enclosures of the range of solutions to initial value problems. Special techniques and software packages have been developed to mitigate this effect.

Because of subdistributivity (4), interval arithmetic has less algebraic structure than most systems of objects commonly used in numerical work, such as real numbers, complex numbers, vectors, and matrices. Besides the extended interval arithmetic described earlier in this section, extensions, such as Kaucher arithmetic, have been constructed in such a way that operations on intervals have a more algebraic structure. The IEEE 1788-2015 interval arithmetic standard has separate accommodation for such systems. However, interpretation of the results in terms of sets and bounds on ranges becomes more complicated in these systems, and we do not describe them further here.

Current Uses

Computation of alpha-cuts in fuzzy information processing is done with interval arithmetic; a good explanation of the connection between these subjects can be found in Kreinovich (2008).

Interval arithmetic is used in various commercial and research global optimization packages to compute bounds on ranges, in constraint propagation algorithms, and in iterative techniques, such as interval Newton methods and other contraction-mapping like iterative techniques. A prominent current commercial package that, although not based on interval arithmetic, uses it effectively in appropriate places is BARON Sahinidis (1996).

More generally, when a system of equations can be posed as a fixed point equation $x = g(x)$, x and $g(x)$ vectors, bounding the range of g with interval arithmetic enables proof of existence and uniqueness of solutions of $x = g(x)$ (or of a related system $f(x) = 0$) within an interval vector $\mathbf{x}$; as such, interval Newton methods lead to an iteration in which the widths of the components of the interval vector iterates $\mathbf{x}_i$ converge quadratically to 0, and each $\mathbf{x}_i$ contains the unique solution to $f(x) = 0$ within the initial interval $\mathbf{x}_0$.

There are prominent specific end applications that effectively use interval arithmetic as part of a strategy to reduce the possibility of failure of systems and to provide certainty in error bounds

and computational results. For instance, interval arithmetic is an important component of systems for automated robotics and automated underwater mapping. Luc Jaulin and colleagues have developed; several of their numerous publications are Jaulin et al. (2001); Jaulin (2009); Desrochers and Jaulin (2017).

Providing guaranteed enclosures to solutions of initial value problems is tricky, given significant uncertainties in the initial conditions and parameters. However, Berz and Makino have successfully incorporated interval arithmetic or related enclosure techniques to provide both certainty and success in various versions of their COSY system for modeling particle accelerator beams Berz et al. (1996).

There is an ongoing list of successes in using interval arithmetic in proving longstanding mathematical conjectures, proving existence and uniqueness in important models and systems of algebraic and differential equations, and the like. Examples of this include proof of the Kepler conjecture Hales (2005); Szpiro (2003), proof that the Lorenz equations have a strange attractor Tucker (1999, 2002), global attractivity of a solution for a population genetics model Bnhelyi et al. (2014), among others.

Reference Resources

We have judged the following references to be the most prominent.

General and Introductory Books

A primary reference, particularly for beginners, is our update Moore et al. (2009) of the first two editions of Ramon Moore's original monograph. The exposition in Moore et al. (2009) is largely on a tutorial level, and Matlab programs are included to illustrate the basic algorithms. However, as in general elementary numerical analysis texts, such programs and algorithms are not replacements for more sophisticated software packages.

Other introductory monographs, written at a higher level, are the original Alefeld and Herzberger work Alefeld and Herzberger (1974) and its English translation Alefeld and Herzberger

(1983), the book Kulisch and Miranker (1981) with a somewhat more specialized point of view, Arnold Neumaier's book Neumaier (1990) with an emphasis on nonlinear algebraic systems of equations and theory, among others. A more recent introductory text, comprehensive and with included mathematical background, available electronically, is Mayer (10 Apr. 2017).

A highly readable and entertaining recent book, requiring only a modest level of prior mathematical knowledge, but with an original and controversial point of view, is Gustafson (2015).

A large number of books containing refereed conference proceedings have also been published.

Applications

The monograph Jaulin et al. (2001) introduces interval methods and develops mathematical structures underlying arguably the most successful applications, such as robotics.

Interval analysis plays an important role in certain branch and bound algorithms for non-convex global optimization of continuous functions. Notable works introducing interval analysis in the context of such algorithms are the second edition Hansen and Walster (2003) of the monograph Hansen (1983) and Kearfott (1996).

Software

Numerous software packages implementing interval arithmetic and associated algorithms have been put forward throughout the years, including packages in specially designed languages and precompilers, as well as in C/C++, Fortran, Python, and others. Some of these are now obsolete or seldom used, while some are new or in current development. The most enduring and popular package presently is arguably INTLAB, a Matlab toolbox Rump (1999); no longer free, it is available for a modest fee. Mathematica has an interval arithmetic package, although it may not have been maintained for some time.

An interesting new package, written in Julia, a Python-like language with special techniques to exploit both the efficiencies of compiled code with the flexibilities of interpreted code, is Benet and Sanders (2020).

Other Resources

Vladik Kreinovich's interval website Kreinovich (2020) lists a huge number of resources and links, such as references, home pages of researchers, software packages, and conferences. The list, although not as selected or curated, is much more comprehensive than given here, and is kept reasonably current.

The site Kearfott (2020) contains a collection of early papers of Ramon Moore and several others.

The *Reliable Computing* journal, originally titled *Interval Computations*, is devoted specifically to interval computations and has been in publication in one form or another and with various publishers since 1991. It is currently an open, refereed online journal published with volunteer work within the interval computations community; see Institute for New Technologies (1991–present) for further information.

The Wikipedia article Wikipedia (2020) has some interesting illustrations and details.

Future Directions

Basic understanding of interval arithmetic is now somewhat mature among experts, but trends can be seen. Of major importance is obtaining a better understanding of what interval arithmetic techniques can and cannot do, along with where and how to use them. There is significantly more such understanding than several decades ago, advances, including demonstrations of important new applications, along with wider knowledge of when interval arithmetic is likely to fail, will continue.

Along with this, the number of people with knowledge of interval arithmetic techniques has grown and will continue to grow. People with knowledge of interval arithmetic were once a close-knit community of experts, but such experts no longer know everyone using the techniques. This is both a blessing and a curse: advances in applications are more likely to occur, but also naive use could lead to frustration and misunderstanding. (“A little knowledge is a dangerous thing.”)

References

- Adams E, Kulisch U (eds) (1993) Scientific computing with automatic result verification, mathematics in science and engineering, vol 189. Academic Press Inc., New York
- Alefeld G, Frommer A, Lang B (eds) (1996) Scientific computing and validated numerics: proceedings of the international symposium on scientific computing, computer arithmetic and validated numerics SCAN-95 held in Wuppertal, Germany, September 26–29, 1995, mathematical research = Mathematische Forschung: Wissenschaftliche Beiträge herausgegeben von der Akademie der Wissenschaften der DDR, Zentralinstitut für Mathematik und Mechanik, vol 90. Akademie-Verlag, Berlin
- Alefeld G, Grigorieff RD (eds) (1980) Fundamentals of numerical computation (computer-oriented numerical analysis), computing supplementum, vol 2. Springer Verlag, Wien/New York
- Alefeld G, Herzberger J (1974) Einführung in die Intervallrechnung. Springer-Verlag, Berlin/Heidelberg/London/etc
- Alefeld G, Herzberger J (1983) Introduction to interval computations. Academic Press Inc., New York, transl. by J. Rokne from the original German ‘Einführung In Die Intervallrechnung’
- Alefeld G, Herzberger J (eds) (1996) Numerical methods and error bounds: proceedings of the IMACS GAMM international symposium on numerical methods and error bounds held in Oldenburg, Germany, July 9–12, 1995, mathematical research, vol 89. Akademie-Verlag, Berlin
- Alt R, Frommer A, Kearfott RB, Luther W (eds) (2004) Numerical software with result verification: International Dagstuhl Seminar, Dagstuhl Castle, Germany, January 19–24, 2003. Revised papers, lecture notes in computer science, vol 2991, Springer-Verlag, Berlin/Heidelberg/London/etc., URL <http://www.springeronline.com/3-540-21260-4>
- Apostolatos N, Kulisch U (1967a) Approximation der Erweiterten Intervallarithmetik Durch die einfache Maschinenintervallarithmetik. Computing 2:181–194
- Apostolatos N, Kulisch U (1967b) Grundlagen einer Maschinenintervallarithmetik. Computing 2:89–104
- Atanassova L, Herzberger J (eds) (1992) Computer arithmetic and enclosure methods: proceedings of the third international IMACS-GAMM symposium on computer arithmetic and scientific computing (SCAN-91), Oldenburg, Germany, 1–4 October 1991, North-Holland. The Netherlands, Amsterdam
- Benet L, Sanders DP (2020) JuliaIntervals web page. Online, open, URL <https://github.com/JuliaIntervals/ValidatedNumerics.jl>
- Berz M, Makino K, Shamseddine K, Hoffstätter GH, Wan W (1996) COSY INFINITY and its applications to nonlinear dynamics. In: Berz M, Bischof C, Corliss G, Griewank A (eds) Computational differentiation: techniques, applications, and tools. SIAM, Philadelphia, pp 363–365
- Bnhelyi B, Csendes T, Krisztin T, Neumaier A (2014) Global attractivity of the zero solution for Wright’s equation. SIAM J Appl Dyn Syst 13(1):537–563. <https://doi.org/10.1137/120904226>
- Burat E (1885) Traité D’Arithmétique a L’usage des élèves des lycées et collèges. Eugène Belin et Fils., Paris, found November 16, 2020 at <https://gallica.bnf.fr/ark:/12148/bpt6k202612w.texteImage>
- Ceberio M, Kreinovich V (eds) (2008) SCAN ‘08: proceedings of the 13th IMACS-GAMM international symposium on scientific computing, computer arithmetic and validated numerics, reliable computing 15 (journal special volume), USA, <http://www.cs.utep.edu/interval-comp/rc.html> or <https://interval.louisiana.edu/reliable-computing-journal/tables-of-contents.html>. Accessed on 28 Nov 2020
- Corliss GF, Kearfott RB (eds) (1993) Abstracts for an international conference on numerical analysis with automatic result verification: mathematics, application and software, February 25–March 1, 1993, Lafayette, LA, 1993, available at <https://interval.louisiana.edu/NAARV-93-abstracts.pdf>. Last accessed 16 Dec 2020
- Csendes T (ed) (1999) Developments in reliable computing: papers presented at the international symposium on scientific computing, computer arithmetic, and validated Numerics, SCAN-98, in Szeged, Hungary, reliable computing, vol 5(3), Kluwer Academic Publishers Group, Norwell/Dordrecht
- Desrochers B, Jaulin L (2017) Computing a guaranteed approximation of the zone explored by a robot. IEEE Trans Autom Control 62(1):425–430. <https://doi.org/10.1109/TAC.2016.2530719>
- Dwyer PS (1951) Computation with approximate numbers. In: Dwyer PS (ed) Linear computations. Wiley & Sons Inc, New York, pp 11–35
- Gauss C (1809) Theoria motus corporum coelestium in sectionibus conicis solem ambientium. Carl Friedrich Gauss Werke, Hamburgi sumptibus Frid. Perthes et I.H. Besser, URL <https://books.google.com/books?id=ORUOAAAAQAAJ>, eBook and PDF currently available free of charge from Google at [https://books.google.com/books/about/Theoria_motus_corporum_coelestium_in_sec.html?id=unhbox\voidb@x\hbox{\\$=\\$}ORUOAAAAQAAJ](https://books.google.com/books/about/Theoria_motus_corporum_coelestium_in_sec.html?id=unhbox\voidb@x\hbox{$=$}ORUOAAAAQAAJ)
- Gustafson J (2015) The end of error: Unum computing. CRC Press/Chapman and Hall, Boca Raton. <https://doi.org/10.1201/9781315161532>
- Hales TC (2005) A proof of the Kepler conjecture. Ann Math 162(3):1065–1185. URL <https://annals.math.princeton.edu/wp-content/uploads/annals-v162-n3-p01.pdf>, together with Samuel P. Ferguson, winner of the 2007 Robbins prize
- Hansen E (ed) (1969) Symposium on interval analysis (1968: Culham, England): topics in interval analysis. Clarendon Press, Oxford, UK
- Hansen E, Walster GW (2003) Global optimization using interval analysis. Marcel Dekker, Inc., New York

- Hansen ER (1983) Global optimization using interval analysis. Marcel Dekker, New York
- Henrici P (1971) Circular arithmetic and the determination of polynomial zeros. In: Conference on application of numerical analysis. Lecture notes in mathematics, vol 228. Springer Verlag, Berlin/Heidelberg/New York, pp 86–92
- Herzberger J (ed) (1994) Topics in validated computations: proceedings of IMACS-GAMM international workshop on validated computation, Oldenburg, Germany, 30 august–3 September 1993, studies in computational mathematics, vol 5. Elsevier, Amsterdam
- IEEE (2015) 1788–2015 – IEEE standard for interval arithmetic. IEEE, New York, <https://doi.org/10.1109/IEEESTD.2015.7140721>, URL <http://ieeexplore.ieee.org/servlet/opac?punumber=7140719>. Approved 11 June 2015 by IEEE-SA Standards Board
- IEEE (2017) IEEE Std 1788.1-2017 – IEEE standard for interval arithmetic (simplified). IEEE, New York, URL <https://standards.ieee.org/findstds/standard/1788.1-2017.html>
- IEEE-754 (2019) IEEE 754-2019, standard for floating-point arithmetic. IEEE, New York. <https://doi.org/10.1109/IEEESTD.2019.876622>
- Institute for New Technologies (1991-present) Reliable Computing journal. <https://interval.louisiana.edu/reliable-computing-journal/RC.html> and <http://www.cs.utep.edu/interval-comp/rcjournal.html>. Accessed 3 Dec 2020
- Jaulin L (2009) Robust set-membership state estimation; application to underwater robotics. Autom 45(1): 202–206. <https://doi.org/10.1016/j.automatica.2008.06.013>
- Jaulin L, Keiffer M, Didrit O, Walter E (2001) Applied interval analysis. SIAM, Philadelphia
- Kaucher EW, Markov SM, Mayer G (eds) (1991) Computer arithmetic, scientific computation and mathematical modelling: proceedings of the second international conference on computer arithmetic, scientific computation and mathematical modelling, Albena, Bulgaria, September 24–28, 1990, IMACS annals on computing and applied mathematics, vol 12. J. C. Baltzer AG, Scientific Publishing Company, Basel
- Kearfott RB (1996) Rigorous global search: continuous problems. No. 13 in nonconvex optimization and its applications. Kluwer Academic Publishers, Dordrecht
- Kearfott RB (2020) Ramon Moore's early papers. Online, open, URL https://interval.louisiana.edu/Moores_early_papers/bibliography.html. Accessed 3 Dec 2020
- Kearfott RB, Kreinovich V (eds) (1996) Applications of interval computations: papers presented at an international workshop in El Paso, Texas, February 23–25, 1995, applied optimization, vol 3. Kluwer Academic Publishers Group, Norwell/Dordrecht
- Knowles S, McAllister WH (eds) (1995) Proceedings of the 12th symposium on computer arithmetic, July 19–21, 1995, Bath, England, symposium on computer arithmetic, vol 12. IEEE Computer Society Press, Silver Spring
- Krämer W, von Gudenberg JW (eds) (2001) Scientific computing, validated numerics, interval methods, Kluwer Academic Publishers Group, Norwell/Dordrecht., URL <http://www.wkap.nl/prod/b/0-306-46706-2>, sCAN 2000, the GAMM–IMACS International symposium on scientific computing, computer arithmetic, and validated numerics and interval 2000, the international conference on interval methods in science and engineering were jointly held in Karlsruhe, September 19–22, 2000
- Kreinovich V (2008) Relations between interval and soft computing. In: Hu C, Kearfott RB, de Korvin A (eds) Knowledge processing with interval and soft computing, Advanced information and knowledge processing. Springer-Verlag, Berlin/Heidelberg/London/etc, pp 75–97. <https://doi.org/10.1007/BFb0085718>
- Kreinovich V (2020) UTEP interval computations web site. Online, open, URL <http://www.cs.utep.edu/interval-comp/main.html>. Accessed 3 Dec 2020
- Kreinovich V, Tucker W (eds) (2012) SCAN '08: Proceedings of the 17th IMACS-GAMM international symposium on scientific computing, computer arithmetic and validated numerics, Reliable Computing 25 (journal special volume), USA, <http://www.cs.utep.edu/interval-comp/rc.html> or <https://interval.louisiana.edu/reliable-computing-journal/tables-of-contents.html>. Accessed 28 Nov 2020
- Kulisch U, Miranker WL (eds) (1983) A new approach to scientific computation: proceedings of a symposium held at IBM Thomas J. Watson Research Center, Yorktown Heights, New York. Academic Press, New York
- Kulisch U, Stetter HJ (eds) (1988) Scientific computation with automatic result verification, computing. Supplementum, vol 6, Springer, Wien/New York, based on papers presented at the conference on computer arithmetic and scientific computation held Sep. 30–Oct. 2, 1987 in Karlsruhe, FRG
- Kulisch UW, Miranker WL (1981) Computer arithmetic in theory and practice. Computer science and applied mathematics. Academic Press Inc., New York
- Luther W, Otten W (eds) (2006) SCAN 2006: proceedings of the 12th GAMM - IMACS international symposium on scientific computing, computer arithmetic and validated Numerics, IEEE Computer Society, USA., <https://www.computer.org/csdl/proceedings/scan/2006/12OmNwwd2WZ>. Accessed 28 Nov 2020
- Mayer G (2017) Interval analysis: and automatic result verification. De Gruyter, Berlin, Boston. <https://doi.org/10.1515/9783110499469>, URL <https://www.degruyter.com/view/title/523499>
- Miranker WL, Toupin RA (eds) (1985) Accurate scientific computations: symposium, Bad Neuenahr, FRG, March 12–14, 1985: proceedings, lecture notes in computer science, vol 235. Springer-Verlag, Berlin/Heidelberg/London/etc
- Moore RE (1959) Automatic error analysis in digital computation. Technical Report Space Div. Report LMSD84821, Lockheed Missiles and Space Co.,

- Sunnyvale, CA, USA, currently available at http://interval.louisiana.edu/Moores_early_papers/Moore_Lockheed.pdf
- Moore RE (1962) Interval arithmetic and automatic error analysis in digital computing. Ph.D. dissertation, Department of Mathematics, Stanford University, Stanford, CA, USA, URL http://interval.louisiana.edu/Moores_early_papers/disert.pdf, also published as applied mathematics and statistics laboratories technical report no. 25
- Moore RE (1966) Interval analysis. Prentice-Hall, Upper Saddle River
- Moore RE (1979) Methods and applications of interval analysis. SIAM, Philadelphia
- Moore RE, Kearfott RB, Cloud MJ (2009) Introduction to interval analysis. SIAM, Philadelphia, URL: <http://www.loc.gov/catdir/enhancements/fy0906/2008042348-b.html>, <http://www.loc.gov/catdir/enhancements/fy0906/2008042348-d.html>; <http://www.loc.gov/catdir/enhancements/fy0906/2008042348-t.html>
- Nehmeier M, von Gudenberg JW, Tucker W (eds) (2016) Scientific computing, computer arithmetic, and validated numerics – 16th international symposium, SCAN 2014, Würzburg, Germany, September 21–26, 2014. Revised selected papers, lecture notes in computer science, vol 9553, Springer, <https://doi.org/10.1007/978-3-319-31769-4>
- Neumaier A (1990) Interval methods for Systems of equations, encyclopedia of mathematics and its applications, vol 37. Cambridge University Press, Cambridge, UK
- Nickel K (1966) Über die Notwendigkeit einer Fehlerschranken- Arithmetik für Rechenautomaten. Numer Math 9:69–79
- Nickel K (1973) The contraction mapping fixed point theorem in interval analysis. MRC technical summary 1334. University of Wisconsin, Madison
- Nickel K (ed) (1975) Interval mathematics: Proceedings of the International Symposium, Karlsruhe, West Germany, May 20–24, 1975, no. 29 in Lecture Notes In Computer Science, Springer Verlag, Berlin/Heidelberg
- Nickel K (ed) (1980) Interval mathematics 1980: proceedings of an international symposium on interval mathematics, held at the Institut für Angewandte Mathematik, Universität Freiburg i. Br., Germany, May 27–31, 1980. Academic Press Inc., New York
- Nickel K (ed) (1986) Interval mathematics 1985: proceedings of the international symposium, Freiburg i. Br., Federal Republic of Germany, September 23–26, 1985. Lecture notes in computer science, vol 212, Springer-Verlag, Berlin/Heidelberg/London/etc
- Rao TRN, Matula DW (eds) (1975) 3rd symposium on computer arithmetic, November 19–20, 1975, Southern Methodist University, Dallas, Texas, IEEE Computer Society Press, Silver Spring, IEEE Cat. No 75 CH1017-3C
- Rohn J (1981) Strong solvability of interval linear programming problems. Computing 26:79–82
- Rump SM (1981) Kleine, exakte Fehlerschranken für die Lösung linearer Gleichungssysteme. Z Angew Math Mech 61:T313–T315
- Rump SM (1999) INTLAB–INTERVAL LABoratory. In: Csendes (1999), pp 77–104, uRL: <http://www.ti3.tu-harburg.de/rump/intlab/>
- Sahinidis NV (1996) BARON: a general purpose global optimization software package. J Glob Optim 8(2): 201–205
- Schnellinger J (1879) Das gekürzte Rechnen (basically, truncated arithmetic or inexact arithmetic). Zeitschrift für das Realschulwesen 1879:257–284, (currently available in Google books)
- Shary S, Corliss G (eds) (2012) SCAN '08: Proceedings of the 15th IMACS-GAMM International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics, Reliable Computing 19 (journal special volume), USA, <http://www.cs.utep.edu/interval-comp/rc.html> or <https://interval.louisiana.edu/reliable-computing-journal/tables-of-contents.html>. Accessed on 28 Nov 2020
- Sunaga T (1958) Theory of interval algebra and its application to numerical analysis. RAAG memoirs 2:29–46, URL <http://www.cs.utep.edu/interval-comp/sunaga.pdf>
- Szpiro GG (2003) Kepler's conjecture: how some of the greatest minds in history helped solve one of the oldest math problems in the world. Wiley
- Tucker W (1999) The Lorenz attractor exists. C R Acad Sci Paris 328:1197–1202
- Tucker W (2002) A rigorous ODE solver and Smale's 14th problem. Found Comput Math 24:53–117
- Warmus MM, Steinhaus H (1956) Calculus of approximations. Bulletin de l'Academie Polonaise des Sciences, CI III IV(5):253–259
- Wikipedia (2020) Interval arithmetic. Online, open, URL https://en.wikipedia.org/wiki/Interval_arithmetic. Accessed 20 Dec 2020
- Young RC (1931) The algebra of many-valued quantities. Math Ann 104:260–290

Aggregation Operators and Soft Computing

Vicenç Torra

Institut d'Investigació en Intel·ligència Artificial –
CSIC, Bellaterra, Spain

Article Outline

Glossary

Definition of the Subject

Introduction

Applications

Aggregation, Integration, and Fusion

Aggregation Operators and Their Construction

Some Aggregation Operators

Future Directions

Bibliography

Glossary

Information Integration The whole process of obtaining some information from different sources and then using this information to achieve a concrete task.

Information Fusion A general term that defines the whole area that studies techniques and methods to combine information for achieving a particular goal.

Aggregation Operators Particular operators that are actually used for combining the information.

Definition of the Subject

Aggregation operators are the particular functions used for combining information in systems where several information sources have to be taken into consideration for achieving a particular goal.

Formally, aggregation operators are the particular techniques of information fusion which can be mathematically expressed in a relatively simple

way. They are part of the more general process of information integration. That is the process that goes from the acquisition of data to the accomplishment of the final task.

As current systems (from search engines to vision and robotics systems) need to consider information from different sources, aggregation operators are of major importance. Such operators permit one to reduce the quantity of information and improve its quality.

Introduction

Information fusion is a broad field that studies computational methods and techniques for combining data and information. At present, different methods have been developed. Differences are on the type of data available and on the assumptions data and information sources (data suppliers) satisfy. For example, methods exist for combining numerical data (e.g., readings from sensors) as well as for combining structured information (e.g., partial plans obtained from AI planners). Besides, the complexity of the methods ranges from simple functions (as the arithmetic mean or the voting technique) that combine a few data to more complex ones (as procedures for dealing with satellite images) that combine a huge burden of data.

Soft computing is an area in computer science that encompasses a few techniques with the goal of solving complex problems when there are no analytic or complete solutions. Two typical problems are the control of a variable (when standard methods, as PID controllers, are not applicable) and the optimization of a function (when standard methods, as quadratic programming, cannot be applied because the objective function to be optimized is highly nonlinear). Among the techniques in soft computing, we can underline genetic algorithms and evolutionary computation, neural networks, and fuzzy sets and fuzzy systems (Sugeno 1974).

Some of the techniques included in soft computing have been used for combining and fusing data. This is the case of neural networks. It is well

known that neural networks (when applied in a supervised manner) can be used to build a system that reproduces up to some extent the behavior of a set of examples. So, in the case of aggregation and data fusion, they can be used to build models when we have a data set consisting on some examples (pairs of input data, expected outcome) that are to be reproduced by the system. Aggregation operators, which can also be seen from the soft computing perspective, follow a different principle. They are defined mathematically and are characterized in terms of their properties.

Aggregation operators are often studied in the framework of fuzzy sets and fuzzy systems, one of the components of soft computing. Such operators, when operating on the $[0,1]$ interval are closely related to t -norms and t -conorms, the basic functions in fuzzy logic to model conjunction and disjunction. Formally, t -norms and t -conorms are used in fuzzy logic to combine the membership degrees either in a conjunctive or disjunctive way. In this setting, aggregation operators are also used to combine information, but additional flexibility is permitted in the combination. Weighted minimum and the fuzzy integrals are some examples of aggregation operators that are typically studied in this framework.

In this entry, we present an overview of the area of information fusion and we focus on the aggregation operators. We review a few of them with some of their properties. We focus on those that have a tight link with soft computing, and especially on the ones that belong to the field of fuzzy sets and fuzzy logic.

Applications

Aggregation operators are used in a large variety of contexts and applications (Bullen 2003). We underline a few of them below. For example, there are applications in economy (to define all kinds of indices, e.g., the retail price index), in biology (to combine DNA sequences and classifications, either expressed as dendrograms – trees – or partitions (Aczél 1966; Fodor and Roubens 1994)), in education (to rate students, universities, and educational systems (Bullen 2003)), in bibliometrics

(to evaluate the quality of the researchers and rate them by means of citation indices, e.g., the number of citations or the h index (Narukawa and Torra 2004; Torra and Narukawa 2008a)), and in computer science.

In computer science, aggregation operators are used for evaluating hardware and software, and for fusing preferences. In the first case, different criteria or performance measures are combined to define an index. For example, performance measures for evaluating runtime systems are defined as the aggregation of the execution times of different benchmarks. Some means (e.g., geometric means) are used for such aggregation. In the second case, different sets of preferences are fused into a new aggregated one. For example, metasearch engines proceed in this way aggregating the results of different search engines. Note that the result of each search engine is a preference on the links and the final ranking is another preference that is built as an aggregation of the initial preferences.

In addition, aggregation operators and information fusion are also used in artificial intelligence, robotics, and vision. Knowledge integration, knowledge revision, or belief merging for knowledge-based systems can also be seen as a case of information fusion at large. In this case, a knowledge base, which is defined using, e.g., description logic in terms of propositions, is checked to detect incoherences and, eventually, to correct them. In vision, fusion is used, e.g., for improving the quality of images. Fusion is also used in robotics for a similar purpose. In particular, it is used for improving the quality of the information obtained from the sensors (Pappus 1982).

The different uses of information fusion in artificial intelligence and robotics can be classified into two main categories. On the one hand, we have the uses related to making decisions and, on the other hand, the uses related with the obtainment of a better description of the application domain (so that the system has a better understanding).

Decision-making typically corresponds to the process of selecting the best alternative. Difficulties in decision-making are found when each alternative is evaluated using different (conflicting) criteria. This case corresponds to the multicriteria decision-making framework (Grabisch et al.

2000; Ruspini et al. 1998). In such framework, depending on the criterion used in the selection, different alternatives are selected. That is, different Pareto optimal solutions can be selected according to the criteria. One approach to solve this situation is to use aggregation operators to combine the degrees of satisfaction for each criterion into a new aggregated degree of satisfaction (for a new *aggregated criterion*). Then selection uses this aggregated criterion to select the best rated alternative. Another approach consists of considering different preferences (one for each criterion), and then such preferences are aggregated into a new aggregated preference. Then, selection is based on the aggregated preference.

A different case, also related to decision-making, is when the alternative to be selected (e.g., the solution of a problem) is constructed from the other ones. For example, several partial solutions are retrieved or obtained by a set of subsystems (maybe because such solutions have been applied before with success into similar problems, as in case-based reasoning). Then, if none of them is completely appropriate in the new situation, the new solution is built from such partial solutions. Some applications of case-based reasoning can be seen from this perspective. This is also the case in plan merging, where partial plans are combined to build the final plan. Another case corresponds to the ensemble methods in machine learning. Such methods assign an outcome (e.g., a class in a classification problem) for a given data taking into account the outcomes of different subsystems (e.g., different classifiers based on different assumptions constructed from the data in the learning set). Then, a final outcome is decided on the basis of the outcomes of the subsystems (e.g., using a voting strategy).

The second use of information fusion and aggregation operators is for improving the understanding of the application domain. This tries to solve some of the problems related to data of low quality, e.g., data that lacks accuracy due to errors caused by the information source or due to the errors introduced in the transmission. Another difficulty is that the data from a single source might describe only part of the application's

domain, and thus, several sources are needed to have a more comprehensive description of the domain, e.g., either several video cameras or a sequence of images from the same video camera is needed to have a full description of a room or a building. Knowledge integration and knowledge revision can be seen from this perspective as the knowledge in a system needs to be updated and refined when the system is under execution.

The different uses of aggregation operators and the large number of applications have stimulated the definition, study, and characterization of aggregation operators on different types of data. Due to this, there are nowadays operators for a large number of different domains. For example, there are operators to combine data in nominal scales (no ordering or structure exists between the elements to be aggregated), ordinal scales (there is ordering between pairs of categories, but no other operators are allowed), numerical data (ratio and measure scales), partitions (input data correspond to a partition of the elements of a given set), fuzzy partitions (as in the previous case, but fuzzy partitions are considered), dendrograms (e.g., tree-like structures for classifying species), DNA sequences (i.e., sequences of A, T, C, and G), images (in remote-sensing applications), and logical propositions (in knowledge integration).

Aggregation, Integration, and Fusion

In the context of information fusion, there are three terms that are in common use. The terms are information integration, information fusion, and aggregation operator. Although these terms are often used as synonymous, their meaning is not exactly equivalent. We define them below.

Information integration	This corresponds to the whole process of obtaining some information from different sources and then using this information to achieve a concrete task. It is also possible that instead of information from several sources, information is extracted from the same source but at different times.
-------------------------	--

(continued)

Information fusion	This is a general term that defines the whole area that studies techniques and methods to combine information for achieving a particular goal. It is also understood as the actual process of combining the information (a particular algorithm that uses the data to obtain a single datum).
Aggregation operators	They are some of the particular operators that are actually used for combining the information. They can be understood as particular information fusion techniques. In fact, such operators are the ones that can be formalized straightforwardly in a mathematical way. The most well-known aggregation operators are the means. Other operators as t -norms and t -conorms, already pointed out above, are also considered by some as aggregation operators. In general, we consider operators that combine N values in a given domain D obtaining a new single datum in the same domain D . For example, in the case of the arithmetic mean, we combine N values on the real line and obtain a new datum that is also a value in the real line.

Taking all this into account, we would say that a robot that combines sensory information includes some information integration processes in its internal system. When sensory data is combined with a mean, or a similar function, we would say that it uses a fusion method that is an aggregation operator. In contrast, if the robot combines the data using a neural network or another black-box approach, we would say that it uses an information fusion method but that no aggregation operator can be distinguished.

As said above, some people distinguish between aggregation operators and means. In this case, a mean is a particular aggregation operator that satisfies unanimity and monotonicity. Other operators that combine N values but that do neither satisfy unanimity nor monotonicity are not means but aggregation operators. Besides, there are aggregation operators that are not required to satisfy monotonicity because they are not defined on ordered domains. This is the case

of operators on nominal scales. The plurality rule or voting is one of such operators.

In this entry, we consider aggregation in a narrow sense. Therefore, an aggregation operator is a function $\mathbb{C}$, from consensus, such that combines N values in a domain D and returns another value in the same domain, i.e., $\mathbb{C} : D^N \rightarrow D$. Such function satisfies unanimity (i.e., $\mathbb{C}(a, \dots, a) = a$ for all a in D) and, if applicable, monotonicity (i. e., $\mathbb{C}(a_1, \dots, a_N) \leq \mathbb{C}(b_1, \dots, b_N)$ when $a_i < b_i$, i.e., this latter condition is valid, as stated before, only when there is an order on the elements in D , and, otherwise, the condition is removed).

Operators on ordered domains defined in this way satisfy internality. That is, the aggregation of values $a_1, \dots, a_N$ is a value between the minimum and the maximum of these values:

$$\min_i a_i \leq \mathbb{C}(a_1, \dots, a_N) \leq \max_i a_i.$$

Aggregation Operators and Their Construction

The definition of an aggregation operator has to take into account several aspects. On the one hand, it should consider the type of data to be combined (this aspect has been mentioned before). That is, whether data is numerical, categorical, etc. On the other hand, it should consider the level of abstraction. That is, whether aggregation has to be done at a lower level or at a higher one. Due to the data flow, the data in a system typically moves from low levels to higher ones. Then, the fusion of two data can be done at different abstraction levels. Note that data at different levels have different representations, e.g., information at a lower level might be numerical (e.g., the values of a measurement), while at a higher level it might be symbolical (e.g., propositions in logic, or belief measures). The level of abstraction and the type of data that are appropriate for fusion depend on a third element, the type of information. That is, information from different sources might be redundant or complementary.

Redundant information is usually aggregated at a low level (e.g., the values of measuring the same object using two similar sensors), while complementary information is usually aggregated at a higher level (e.g., the fusion of images to have a better representation of 3D objects cannot be done at a pixel-signal-level).

Once the appropriate environment for a function has been settled, it is time for the definition of the function. Two main cases can be distinguished, (i) definition from examples and (ii) definition from properties.

1. The use of examples for defining the operators follows machine learning theory. Two main cases can be distinguished. The first one is when we assume that the model of aggregation is known, e.g., we know that we will use a weighted mean for the aggregation, and then the problem is to find the appropriate parameters for such weighted mean. The second one is when the model is not known and only a loose model – black box model – is assumed. Then the examples are used to settle such model. The use of neural networks is an example of this kind of application.
2. In the case of definition from properties, the system designer establishes the properties that the aggregation operator needs to satisfy. Then the operator is derived from such properties.

Functional equations (Aczél 1987; Arnold et al. 1992) can be used for this purpose. Functional equations are equations that have functions as their unknowns. For example, $\phi(x + y) = \phi(x) + \phi(y)$ is an example of a functional equation (one of Cauchy's equations) with an unknown function ϕ . In this equation, the goal is to find the functions ϕ that satisfy such equation. In this case, when ϕ are continuous, solutions are functions $\phi(x)$ of the form $\phi(x) = ax$.

In the case of aggregation, we might be interested in, for instance, functions $\phi(x, y)$ that aggregate two numerical values x and y , and that satisfy unanimity (i.e., $\phi(x, x) = x$), invariance to transformations of the form $\gamma(x) = x + t$ (i.e., $\phi(x_1 + t, x_2 + t) = \phi(x_1, x_2) + t$), and invariance to transformations of the form $\gamma(x) = rx$ - positive homogeneity (i.e., $\phi(rx_1, rx_2) = r\phi(x_1, x_2)$). Such

functions are completely characterized, and all of them are of the form $\phi(x_1, x_2) = (1 - k)x_1 + kx_2$.

Naturally, the definition from properties might cause the definition of an overconstrained problem. That is, we may include too restrictive constraints and define a problem with no possible solution. This is the case of the problem of aggregation of preferences (Bajraktarević 1958) that lead to Arrow's impossibility theorem (Arrow et al. 2002). That is, there is a set of natural properties that no function for the aggregation of preferences can satisfy.

The definition of the aggregated value as the one that is located at a minimum distance from the objects being aggregated is another approach for defining aggregation operators. In this case, we need to define a distance between pairs of objects on the domain of application. Let D be the domain of application, and let d be a distance function for pairs of objects in D (that is, $d : D \times D \rightarrow \mathbb{R}$). Then, the aggregation function $\mathbb{C} : D^N \rightarrow D$ applied to $a_1, \dots, a_N$ is formally defined as

$$\mathbb{C}(a_1, \dots, a_N) = \arg \min_c \left\{ \sum_{i=1}^N d(c, a_i) \right\}.$$

Such definition is often known as the median rule. In the numerical setting, i.e., if the domain D is a subset of the real set, when $d(a, b) = (a - b)^2$, the aggregation operator results to be the arithmetic mean. When $d(a, b) = |a - b|$, the aggregation operator results into the median. Formally, the median is the only solution when N is of the form $2k + 1$ or when N is of the form $2k$, but the two central values are equal. Otherwise (when N is of the form $2k$ and the two central values are different), the median is one of the solutions of the equation above, but other solutions might be also valid. For example, let us assume that $N = 4$ and that $a_1 \leq a_2 < a_3 \leq a_4$, then another solution is,

$$\mathbb{C}(a_1, \dots, a_N) = \frac{(a_4 a_3 - a_2 a_1)}{(a_4 + a_3) - (a_2 + a_1)}.$$

When $d(a, b) = |a - b|^\infty$ for $k \rightarrow \infty$, we have that the midrange of the elements $a_1, \dots, a_N$ is a solution of the equation above. That is,

$$\mathbb{C}(a_1, \dots, a_N) = \frac{(\min_i a_i - \max_i a_i)}{2}.$$

Some Aggregation Operators

There exist a large number of aggregation operators. We discuss in this section a few of them, classified into different families. We begin with the quasi-arithmetic mean, a family of operators that encompasses some of the most well-known aggregation operators, the arithmetic mean, the geometric mean, and the harmonic mean.

Quasi-Arithmetic Means

A quasi-arithmetic mean is an aggregation operator $\mathbb{C}$ of the form

$$\text{QAM}(a_1, \dots, a_N) = \varphi^{-1} \left(\frac{1}{N} \sum_{i=1}^N \varphi(a_i) \right).$$

with φ an invertible function.

We list below a few examples of quasi-arithmetic means.

- Arithmetic mean:

$$\text{AM}(a_1, \dots, a_N) = \frac{1}{N} \sum_{i=1}^N a_i.$$

- Geometric mean:

$$\text{GM}(a_1, \dots, a_N) = \sqrt[N]{\frac{1}{N} \sum_{i=1}^N a_i}.$$

- Harmonic mean:

$$\text{HM}(a_1, \dots, a_N) = \frac{N}{\sum_{i=1}^N \frac{1}{a_i}}.$$

- Power mean (Hölder mean, or root-mean-power, RMP):

$$\text{RMP}(a_1, \dots, a_N) = \sqrt[\alpha]{\frac{1}{N} \sum_{i=1}^N a_i^\alpha}.$$

- Quadratic mean (root-mean-square):

$$\text{RMS}(a_1, \dots, a_N) = \sqrt{\frac{1}{N} \sum_{i=1}^N a_i^2}.$$

These means are obtained from the quasi-arithmetic mean using the following definition for the function φ : $\varphi(x) = x$ (in the arithmetic mean), $\varphi(x) = \log x$ (in the geometric mean), $\varphi(x) = 1/x$ (in the harmonic mean), $\varphi(x) = x^\alpha$ (in the power mean), and $\varphi(x) = x^2$ (in the quadratic mean).

Several properties have been proved for these functions. In particular, some inequalities have been proven (Choquet 1953; Hirsch 2005). For example, it is known that the harmonic mean (HM) always returns a value that is smaller than the one of the geometric mean (GM) and that the value of the arithmetic mean is smaller than the one of the arithmetic mean (AM). That is, the following inequality holds for all x, y ,

$$\text{HM}(x, y) \leq \text{GM}(x, y) \leq \text{AM}(x, y).$$

In the case of the power means, the following inequality holds,

$$\text{RMP}_r(x, y) \leq \text{RMP}_s(x, y) \quad \text{for all } r < s.$$

That is, the power means is monotonic with respect to its parameter.

Also related with the power means, we have that when $\alpha \rightarrow 0$, $\text{RMP}_\alpha = \text{GM}$, that when $\alpha \rightarrow \infty$, $\text{RMP}_\alpha = \max$ and that when $\alpha \rightarrow -\infty$, $\text{RMP}_\alpha = \min$.

From the point of view of the applications, it is of relevance whether the outcome of an

aggregation operator is, in general, large or small. In other words, whether an aggregation operator A is larger or smaller than another aggregation operator B . For binary operators A and B , this corresponds to know whether it holds or not that

$$A(x, y) \leq B(x, y) \quad \text{for all } x, y.$$

Due to the importance in real practice of such ordering, some indices have been developed for analyzing the outcomes of the operators. The average value, the orness, and the andness are examples of such indices.

The orness computes the similarity of the operator with the maximum. Recall that in fuzzy sets, the maximum is used to model the disjunction, i.e., or. In contrast to that, the andness computes the similarity of the operator with the minimum. Recall that in fuzzy sets, the minimum is used to model the conjunction, i.e., and. For these measures, it can be proven that (i) the operator maximum has a maximum orness (an orness equal to one) and a minimum andness (an andness equal to zero), (ii) the orness of the arithmetic mean is larger than the orness of the geometric mean, and this latter orness is larger than the one of the harmonic mean. This property follows from the following inequality that holds for all x and y ,

$$HM(x, y) \leq GM(x, y) \leq AM(x, y).$$

The interpretations for the andness and the orness ease the selection of the appropriate aggregation function. This is so because the orness can be interpreted as a degree of compensation. That is, when different criteria are considered, a high value of orness means that a few good criteria can compensate a large number of bad ones. In contrast to that, a low value of orness means that all criteria have to be satisfied in order to have a good final score. Taking this into account, and using the example above, when only low compensation is permitted, it is advisable to use the geometric mean or the harmonic mean. In contrast, a larger compensation requires the user to consider the arithmetic mean.

A similar approach can be applied for the power mean. The larger the parameter r in the function (i.e., RMP_r), the larger the compensation is.

Weighted Means and Quasi-Weighted Means

The weighted mean is the most well-known aggregation operator with weights, and the quasi-weighted mean is one of its generalizations. In such operator, weights are represented by means of a weighting vector. That is, a vector $w = (w_1, \dots, w_N)$ that satisfies

$$w_i \geq 0 \quad \text{and} \quad \sum w_i = 1.$$

- Weighted mean: given a weighting vector $w = (w_1, \dots, w_N)$, the weighted mean is defined as follows

$$WM(a_1, \dots, a_N) = \sum_{i=1}^N w_i a_i.$$

- Naturally, when $w_i = 1/N$ for all $i = 1, \dots, N$, the weighted mean corresponds to the arithmetic mean.
- In this definition, the weights can be understood as their reliability. The larger the importance, the larger the weights.
- Quasi-weighted mean: given a weighting vector $w = (w_1, \dots, w_N)$, the quasi-weighted mean (or quasi-linear mean) is defined as follows

$$QWM(a_1, \dots, a_N) = \varphi^{-1} \left(w_i \sum_{i=1}^N \varphi(a_i) \right),$$

- With φ an invertible function, and $w = (w_1, \dots, w_N)$ a weighting vector.
- The quasi-weighted mean generalizes the quasi-arithmetic mean given above introducing weights w_i for each of the source. Using appropriate functions φ , we obtain the weighted mean (with $\varphi(x) = x$), the weighted geometric mean

(with $\varphi(x) = \log x$), the weighted harmonic mean (with $\varphi(x) = 1/x$), the weighted power mean (with $\varphi(x) = x^a$), and the weighted quadratic mean (with $\varphi(x) = x^2$).

Losonczi's Means

This family of means, defined in (Barthelemy and McMorris 1986), is related to the quasi-weighted means. While in the former the weights only depend on the information source (there is one weight for each a_i , but such weight does not depend on the value a_i), this is not the case in the Losonczi's means. In this operator, weights depend on the values being aggregated.

Formally, a Losonczi's mean is an aggregation operator $\mathbb{C}$ of the form,

$$\mathbb{C}(a_1, \dots, a_N) = \varphi^{-1} \left(\frac{\sum_{i=1}^N \pi_i(a_i) \varphi(a_i)}{\sum_{i=1}^N \pi_i(a_i)} \right),$$

with φ an invertible function.

Naturally, when $\pi_i(a)$ is constant for all a , we have that the function reduces to a quasi-weighted means. When $\pi_i(a) = a^q$ and $\varphi(a) = a^{p-q}$ with $q = p - 1$, the Losonczi's means reduces to the counterharmonic mean, also known by Lehmer mean. That is,

$$\mathbb{C}(a_1, \dots, a_N) = \frac{\sum_{i=1}^N a_i^p}{\sum_{i=1}^N a_i^{p-1}}.$$

When $\pi_i(a) = w_i a^q$ and $\varphi(a) = a^{p-q}$ with $q = p - 1$, the Losonczi's means reduces to the weighted Lehmer mean. That is,

$$\mathbb{C}(a_1, \dots, a_N) = \frac{\sum_{i=1}^N w_i a_i^p}{\sum_{i=1}^N w_i a_i^{p-1}}.$$

Linear Combination of Order Statistics and OWA Operators

Given $a_1, \dots, a_N$, the i th order statistics (Arrow 1951) is the element in $a_1, \dots, a_N$ that occupies the i th position when such elements are ordered in increasing order. Formally, let i be an index $i \in \{1, \dots, N\}$. Then, the i th order statistics OS_i of dimension N is defined as follows:

$$OS_i(a_1, \dots, a_N) = a_{s(i)}.$$

Where $\{s(1), \dots, s(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{s(j)} \leq a_{s(j+1)}$ for all $j \in \{1, \dots, N-1\}$.

When N is of the form $2k + 1$, $OS_{(N+1)/2} = OS_{k+1}$ is the median; when N is of the form $2k$, the median corresponds to the mean value of $OS_{N/2}$ and $OS_{N/2+1}$. That is,

$$\text{median}(a_1, \dots, a_N) = (OS_{N/2} + OS_{N/2+1})/2.$$

The OWA (ordered weighted averaging) operator is a linear combination of order statistics (an L -estimator). Its definition is as follows:

- OWA operator: given a weighting vector $w = (w_1, \dots, w_N)$ (i.e., a vector such that $w_i \geq 0$ and $\sum_i w_i = 1$), the OWA operator is defined as follows

$$\text{OWA}(a_1, \dots, a_N) = \sum_{i=1}^N w_i a_{\sigma(i)},$$

where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$. Equivalently,

$$\text{OWA}(a_1, \dots, a_N) = \sum_{i=1}^N w_i \cdot OS_i(a_1, \dots, a_N).$$

Naturally, when $w_i = 1/N$ for all $i = 1, \dots, N$, the OWA operator corresponds to the arithmetic mean.

- The ordering of the elements, permit the user to express the importance of some of the values being aggregated (importance of the values with respect to their position). For example, the OWA permits us to model situations where importance is given to low values (e.g., in a robot, low values are more important than larger ones to avoid collisions). So, informally, we have that the w_i for $i = N$ and $i = N - 1$ should be large.
- Geometric OWA operator: Given a weighting vector $w = (w_1, \dots, w_N)$, the geometric OWA (GOWA) operator is defined as follows

$$\text{GOWA}(a_1, \dots, a_N) = \prod_{i=1}^N a_{\sigma(i)}^{w_i},$$

- Where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$. That is, the GOWA operator is a geometric mean of the order statistics.
- Generalized OWA operator: Given a weighting vector $w = (w_1, \dots, w_N)$, the generalized OWA operator is defined as follows

$$\begin{aligned} \text{Generalized OWA}(a_1, \dots, a_N) \\ = \left(\sum_{i=1}^N w_i a_{\sigma(i)}^\alpha \right)^{1/\alpha}, \end{aligned}$$

where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$. That is, the generalized OWA operator is a generalized mean (root-mean-powers) of the order statistics.

- Quasi-OWA operator: Given a weighting vector $w = (w_1, \dots, w_N)$, the quasi-OWA (QOWA) operator is defined as follows

$$\text{QOWA}(a_1, \dots, a_N) = \varphi^{-1} \left(\sum_{i=1}^N w_i \varphi(a_{\sigma(i)}) \right).$$

where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$. That is, the quasi-OWA operator is a quasi-weighted mean of the order statistics.

$$\text{OWA}(a_1, \dots, a_N) = \prod_{i=1}^N a_{\sigma(i)}^{w_i},$$

- Where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $a_{\sigma(i-1)} \geq a_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$.
- BADD-OWA operator: Given a weighting vector $w = (w_1, \dots, w_N)$, the BADD-OWA (basic defuzzification distribution OWA) corresponds to the counterharmonic mean (see above). That is,

$$\mathbb{C}(a_1, \dots, a_N) = \frac{\sum_{i=1}^N a_i^p}{\sum_{i=1}^N a_i^{p-1}}.$$

- Induced OWA operator: Given a weighting vector $w = (w_1, \dots, w_N)$ and a priority vector $b = (b_1, \dots, b_N)$, the induced OWA (IOWA) operator is defined as follows

$$\text{IOWA}_b(a_1, \dots, a_N) = \sum_{i=1}^N w_i a_{\sigma(i)},$$

- Where $\{\sigma(1), \dots, \sigma(N)\}$ is a permutation of $\{1, \dots, N\}$ such that $b_{\sigma(i-1)} \geq b_{\sigma(i)}$ for all $i \in \{2, \dots, N\}$. Naturally, the OWA corresponds to the IOWA when a is used as the priority vector.

The WOWA Operator

Let p and w be two weighting vectors of dimension N . Then, the WOWA (Torra 2003) operator (weighted ordered weighted averaging operator) of dimension N is defined as

$$\text{WOWA}_{p,w}(a_1, \dots, a_N) = \sum_{i=1}^N \omega_i a_{\sigma(i)},$$

where σ is defined as above (i.e., $a_{\sigma(i)}$ is the i th largest element in the collection $a_1, \dots, a_N$), and the weight ω_i is defined as

$$\omega_i = w^* \left(\sum_{j \leq i} p_{\sigma(j)} \right) - w^* \left(\sum_{j < i} p_{\sigma(j)} \right),$$

with w^* being a nondecreasing function that interpolates the points

$$\left\{ \left(i/N, \sum_{j \leq i} w_j \right) \right\}_{i=1, \dots, N} \cup \{(0, 0)\}.$$

The function w^* is required to be a straight line when the points can be interpolated in this way.

The WOWA operator generalizes the OWA and the weighted mean. When $p = (1/N, \dots, 1/N)$, the WOWA reduces to an OWA operator and when $w = (1/N, \dots, 1/N)$, the WOWA reduces to a weighted mean. That is,

$$\begin{aligned} \text{WOWA}_{p,w}(a_1, \dots, a_N) &= \text{OWA}_w(a_1, \dots, a_N), \\ &\text{when } p = (1/N, \dots, 1/N), \end{aligned}$$

and

$$\begin{aligned} \text{WOWA}_{p,w}(a_1, \dots, a_N) &= \text{WM}_p(a_1, \dots, a_N), \\ &\text{when } w = (1/N, \dots, 1/N). \end{aligned}$$

As the OWA and the weighted mean generalize the arithmetic mean, when the weights are all equal to $1/N$, the same happens with the WOWA, i.e.,

$$\begin{aligned} \text{WOWA}_{p,w}(a_1, \dots, a_N) &= \text{AM}(a_1, \dots, a_N), \\ &\text{when } p = (1/N, \dots, 1/N) \text{ and} \\ &w = (1/N, \dots, 1/N). \end{aligned}$$

As stated here, the WOWA operator reduces to the OWA and the weighted mean for appropriate

weighting vectors. This property permits the user to interpret the two weighting vectors p and w used in the WOWA in terms of the corresponding interpretation in the OWA and the weighted mean. That is, p corresponds to the reliability of the information sources and w permits to express the importance of the values (with respect to their ordering). Thus, the WOWA permits the user to express which are the most reliable information sources (e.g., sensors) and which are the values that have to be considered more relevant (e.g., larger importance to small values in sensors to avoid collisions, or larger importance to large values if a high degree of compensation is permitted in multicriteria decision-making).

Fuzzy Integrals

Fuzzy integrals can be used in aggregation (Hardy et al. 1934; Torra and Narukawa 2008a). Formally, they integrate a function with respect to a fuzzy measure. From the aggregation point of view, the function corresponds to the data to be aggregated and the fuzzy measure corresponds to some background knowledge (following the jargon in artificial intelligence) about the reliability or importance of the information sources. In this sense, the fuzzy measure can be seen as a generalization of the weights as they are used for the same purpose. The difference is that while the weights can only be used for a single information source, the measures are applied to sets of them.

In order to use the fuzzy integrals in aggregation, we need first some formalization. First, we need to define the set of information sources, let it be $X = \{x_1, \dots, x_N\}$. That is, X is the set of sensors in a robot or the set of criteria in a multicriteria decision-making problem. Then, we need to express the datum (a number) supplied by source x_i . This is represented by a function f from X to $\mathbb{R}$. That is, $f(x_i)$ is the data supplied by sensor x_i or the evaluation of the criterion x_i . In terms of the notation used above in this work, $f(x_i) = a_i$. Then the aggregation of the data supplied by the information sources X can be computed as the integral of the function f with respect to a fuzzy measure on X . The measure is a function defined on the subsets of X . So, for any set of information sources, we have its *weight*.

Formally, a fuzzy measure is a set function on X into $[0,1]$. That is, a function from $\wp(X)$ into $[0,1]$ or, in other words, a function that assigns a value in $[0,1]$ to each subset A of X . Furthermore, fuzzy measures are monotonic with respect to set inclusion. Their definition is as follows.

A fuzzy measure μ is a set function $\mu: \wp(X) \rightarrow [0, 1]$ that satisfies:

- $\mu(\emptyset) = 0$,
- $\mu(X) = 1$,
- $A \subseteq B$ implies $\mu(A) \leq \mu(B)$.

When the measure of A is understood as its importance, the last condition in this definition states that the more sources we have, the larger their importance. Boundary conditions (the first two conditions) imply that the minimum importance (the one of the empty set) is zero and that the maximum importance (the one of the full set) is one. The boundary condition $\mu(X) = 1$ is arbitrary; nevertheless, from the point of view of the aggregation, it can be justified as being similar to the one existing for the weights in the weighted mean, and having a similar rationale. That is, the weights w_i in a weighting vector add to one: $\sum_i w_i = 1$. Note that such restriction is appropriate in the case of the weighted mean because it implies that the weighted mean satisfies unanimity (WM $(a, \dots, a) = a$). Note that any other arbitrary bound K (i.e., $\sum_i w_i = K$) would just result into a weighted mean such that unanimity does not hold but that $\text{WM}(a, \dots, a) = Ka$.

Due to the usefulness of the fuzzy measures (Hardy et al., 1934), several families have been defined and studied in the literature. Some of them are the following: Sugeno lambda measures, decomposable fuzzy measures, hierarchically decomposable fuzzy measures, and k -order additive fuzzy measures.

As said, fuzzy measures are used in combination with fuzzy integrals. Several fuzzy integrals have been defined in the literature. We review a few of them below.

- Choquet integral (Fishburn and Rubinstein 1986): The Choquet integral of a function $f: X \rightarrow \mathbb{R}^+$ with respect to a fuzzy measure μ is defined by:

$$\begin{aligned} \text{CI}_\mu(f) &= (C) \int f d\mu \\ &= \sum_{i=1}^N [f(x_{s(i)}) - f(x_{s(i-1)})] \mu(A_{s(i)}). \end{aligned}$$

where $f(x_{s(i)})$ indicates that the indices have been permuted so that $0 \leq f(x_{s(1)}) \leq \dots \leq f(x_{s(N)}) \leq 1$, where $f(x_{s(0)}) = 0$ and where $A_{s(i)} = \{x_{s(i)}, \dots, x_{s(N)}\}$.

The Choquet integral generalizes the arithmetic mean, the weighted mean, the OWA operator, and the WOWA operator.

1. It generalizes the arithmetic mean when we have that the measure is proportional to the cardinality of the set, i.e., when $\mu(A) = |A| / |X|$, the Choquet integral of f corresponds to the arithmetic mean of $a_i = f(x_i)$.
2. Given a weighting vector $w = (w_1, \dots, w_N)$, if we define a fuzzy measure as follows

$$\mu(A) = \sum_{x_i \in A} w_i,$$

3. we have that the Choquet integral of f with respect to μ corresponds to the weighted mean of a_i with respect to w .
4. Given a weighting vector $w = (w_1, \dots, w_N)$, if we define a fuzzy measure μ as follows.

$$\mu(A) = \sum_{i=1}^{|A|} w_i,$$

5. We have that the Choquet integral of f with respect to μ corresponding to the OWA of a_i with respect to w .
6. Given weighting vectors w and p , we can define a fuzzy measure in terms of the interpolating function w^* . Once such function is built, we define the fuzzy measure μ as follows:

$$\mu(A) = w^* \left(\sum_{i=1}^{|A|} w_i \right),$$

7. We have that the Choquet integral of f with respect to such μ corresponds to the WOVA of the a_i with respect to weighting vectors w and p .
- Sugeno integral (Torra 1997): The Sugeno integral of a function $f: X \rightarrow \mathbb{R}^+$ with respect to a fuzzy measure μ is defined by:

$$\begin{aligned} \text{SI}_\mu(f) &= (S) \int f d\mu \\ &= \max_{i=1}^N \left(\min \left(f(x_{s(i)}), \mu(A_{s(i)}) \right) \right). \end{aligned}$$

- Here, s and $A_{s(i)}$ are defined as above.

Other fuzzy integrals exist. Among them, we have the twofold integral that generalizes both Choquet and Sugeno integrals. The generalization uses two fuzzy measures μ_C and μ_S that correspond, respectively, to the measures used in the Choquet and Sugeno integrals. The measure of the

Choquet integral has a probabilistic flavor as the Choquet integral generalizes the weighted mean that can be seen as the expectation of the a_i with respect to the probability distribution $w = (w_1, \dots, w_N)$. Note that the weighting vector in the weighted mean is compatible with a probability distribution because weights are positive and add to one. Instead, the Sugeno integral has a fuzzy flavor as it generalizes the weighted min. and the weighted max., which are used in fuzzy systems. Such operators are defined in terms of weighting vectors that do not add to one but have a maximum of one. So, they are like possibility distributions.

In short, the twofold integral integrates a function with respect to two fuzzy measures μ_C and μ_S . We define such integral below.

- Twofold integral (Mitchell 2007a; Torra 2003): The twofold integral of a function $f: X \rightarrow \mathbb{R}^+$ with respect to a fuzzy measure μ is defined by:

$$\text{TI}_{\mu_S, \mu_C}(f) = \sum_{i=1}^N \left(\left(\max_{j=1}^i \left(\min \left(f(x_{s(i)}), \mu_S(A_{s(i)}) \right) \right) \right) \cdot \left(\mu_C(A_{s(i)}) - \mu_C(A_{s(i+1)}) \right) \right).$$

Here, s and $A_{s(i)}$ are defined as above. Naturally, $A_{s(n+1)} = \emptyset$.

When $\mu_C = \mu^*$, the twofold integral reduces to the Sugeno integral and when $\mu_S = \mu^*$, the twofold integral reduces to the Choquet integral. Here, μ^* represents ignorance and is defined by $\mu(\emptyset) = 0$ and $\mu^*(A) = 1$ when $A \neq \emptyset$. In addition, when $\mu_C = \mu_S = \mu^*$, we have that the twofold integral reduces to the maximum.

Future Directions

Aggregation operators have been studied for a long time. For example, the inequality

$$\text{HM}(x, y) \leq \text{GM}(x, y) \leq \text{AM}(x, y),$$

that involves the harmonic mean (HM), the geometric mean (GM), and the arithmetic mean (AM) was already known by Pappus of Alexandria (fl. c. 300–c. 350). It is given in his book *Synagoge* (book III). Since then, a large number of aggregation operators have been proposed fostered by the new applications and the theoretical studies.

From the practical point of view, current research is oriented toward the embedding of such operators into real applications. This has motivated the study and the development of methods and techniques for parameter determination, e.g., the development of methods that permits a user to find the appropriate parameters in a given

application. For example, methods have been defined to determine the weights in the weighted mean, and to determine the fuzzy measure in a fuzzy integral. This is an active line of research.

Another related topic is the selection of the appropriate function. Now, the methods for parameter determination mainly presume that the appropriate operator is known beforehand and that the only elements to be determined are the parameters. Tools and methods are needed to help the users to select which is the appropriate function.

Finally, there is the need for developing architectures for information integration. That is, architectures that encompass all the processes related with fusion: from the input data (acquisition of the data) to the final decision or output.

All these aspects, oriented toward the real application of aggregation operators, are combined with new research on the theoretical side. Such research includes the definition of new functions (needed so that systems can work on new types of data), the characterization of the functions (needed so that we can know which are the properties of the functions and so that a user can then select the appropriate one based on their problem requirements), and the study of composite models using aggregation operators (i.e., hierarchical models that combine several aggregation operators).

Bibliography

Primary Literature

- Aczél J (1966) Lectures on functional equations and their applications. Academic, New York, London
- Aczél J (1987) A short course on functional equations. Reidel, Dordrecht
- Arnold BC, Balakrishnan N, Nagaraja HN (1992) A first course in order statistics. Wiley, New York
- Arrow KJ (1951) Social choice and individual values, 2nd edn. Wiley, New York
- Arrow KJ, Sen AK, Suzumura K (eds) (2002) Handbook of social choice and welfare. Elsevier, Amsterdam
- Bajraktarević M (1958) Sur une équation fonctionnelle aux valeurs moyennes. Glasnik Mat Fiz I Astr 13(4): 243–248
- Barthelemy JP, McMorris FR (1986) The median procedure for n-trees. J Classif 3:329–334
- Bouyssou D, Marchant T, Pirlot M, Perny P, Tsoukiàs A, Vincke P (2000) Evaluation and decision models: a

- critical perspective. Kluwer's International Series. Kluwer, Dordrecht
- Bullen PS (2003) Handbook of means and their inequalities. Kluwer, Dordrecht
- Bullen PS, Mitrinović DS, Vasić PM (1988) Means and their inequalities. Reidel, Dordrecht
- Choquet G (1953) Theory of capacities Ann Inst Fourier 5: 131–295
- Fishburn PC, Rubinstein A (1986) Aggregation of equivalence relations J Classif 3:61–65
- Fodor J, Roubens M (1994) Fuzzy preference modelling and multicriteria decision support. Kluwer, Dordrecht
- Grabisch M, Murofushi T, Sugeno M (2000) Fuzzy measures and integrals: theory and applications. Physica, Heidelberg
- Hardy GH, Littlewood JE, Pólya G (1934) Inequalities, 2nd edn. Cambridge University Press, Cambridge
- Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci 102(45):16569–16572
- Mitchell HB (2007a) Multi-sensor data fusion. An introduction, Springer, Heidelberg
- Narukawa Y, Torra V (2004) Twofold integral and multi-step Choquet integral. Kybernetika 40(1):39–50
- Pappus (1982) La collection mathématique. Librairie Scientifique et Technique Albert Blanchard, Paris
- Roy B (1996) Multicriteria methodology for decision aiding. Kluwer, Dordrecht
- Ruspini EH, Bonissone PP, Pedrycz W (1998) Handbook of fuzzy computation. IOP, London
- Sugeno M (1974) Theory of fuzzy integrals and its applications. Ph D dissertation. Tokyo Institute of Technology, Japan
- Torra V (1997) The weighted OWA operator. Int J Intell Syst 12:153–166
- Torra V (2003) La integral doble o *twofold* integral: Una generalització de les integrals de Choquet i Sugeno. Butlletí de l'Associació Catalana d'Intel·ligència Artificial 29:13–19. Preliminary version in English: Twofold integral: A generalization of Choquet and Sugeno integral. IIIA Technical Report TR-2003-08
- Torra V, Narukawa Y (2007) Modeling decisions: information fusion and aggregation operators. Springer, Heidelberg
- Torra V, Narukawa Y (2008a) The h-index and the number of citations: two fuzzy integrals. IEEE Trans Fuzzy Syst 16(3):795–797

Books and Reviews

- Alsina C, Frank MJ, Schweizer B (2006) Associative functions: triangular norms and copulas. World Scientific, Singapore
- Calvo T, Mayor G, Mesiar R (2002) Aggregation operators. Physica, Heidelberg
- Mitchell (2007b) is an introduction to multisensor data fusion, a topic very much related with aggregation operators
- Pap E (2002) Handbook of measure theory, vols I, II. North-Holland, Amsterdam

Torra and Narukawa (2008b) gives a general description of the field of aggregation operators, it defines the main operators and discusses a few practical topics about their applications (e. g. parameter determination). (Calvo et al. 2002) is an edited book that contains state-of-the-art chapter on different topics related with aggregation and fusion. A few properties on the aggregation operators (mainly related with inequalities) can

be found in the books by Bullen (2003) and Bullen, Mitrović and Vasić (1988), and the excellent book by Hardy, Littlewood and Pólya (1934). Grabisch et al. (2000) is an edited volume on fuzzy measures and fuzzy integrals

Yager RR (1988) On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans Syst Man Cybern* 18:183–190

Evolving Fuzzy Systems

Plamen Angelov

Intelligent Systems Research Laboratory, Digital
Signal Processing Research Group,
Communication Systems Department, Lancaster
University, Lancaster, UK

Article Outline

Glossary

Definition of the Subject

Introduction

Evolving Clustering

Evolving TS Fuzzy Systems

Evolving Fuzzy Classifiers

Evolving Fuzzy Controllers

Application Case Studies

Future Directions

Bibliography

Glossary

Evolving system In the context of this article the term ‘evolving’ is used in the sense of the self-development of a system (in terms of both its structure and parameters) based on the stream of data coming to the system on-line and in real-time from the environment and the system itself. The system is assumed to be mathematically described by a set of fuzzy rules of the form:

Ruleⁱ:

IF ($Input_1$ is close to prototype^{*i*}₁)

AND ... *AND* ($Input_n$ is close to prototype^{*i*}_{*n*}) (1)

THEN ($Output^i = \overline{Inputs}^T ConseqParams$)

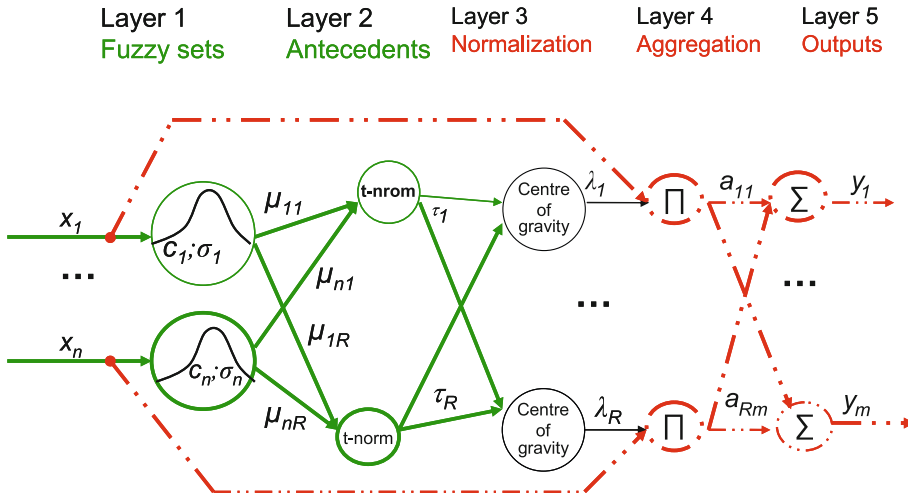
In this sense, this definition strictly follows the meaning of the English word “evolving” as described in (Hornby 1974), p. 294, namely “unfolding; developing; being developed,

naturally and gradually”. Contrast this to the definition of “evolutionary” in the same source, which is “development of more complicated forms of life (plants, animals) from earlier and simpler forms”. The terms evolutionary or genetic are also associated with such phenomena (respectively operators that mimic these) as chromosome crossover, mutation, selection and reproduction, parents and offsprings (Goldberg 1989). Evolving (fuzzy and neuro-fuzzy) systems do not deal with such phenomena. They rather consider a *gradual development* of the underlying (fuzzy or neuro-fuzzy) system structure.

Fuzzy system structure Structure of a fuzzy (or neuro-fuzzy) system is constituted of a set of fuzzy rules (1). Each fuzzy rule is composed of antecedent (IF) and consequents (THEN) parts. They are linguistically expressed. The antecedent part consists of a number of fuzzy sets that are linked with fuzzy logic aggregators such as conjunction, disjunction, more rarely, negation (Klir and Folger 1988). In the above example, a conjunction (logical AND) is used. It can be mathematically described by so-called t-norms or t-conorms between membership functions. The most popular membership functions are Gaussian, triangular, trapezoidal (Yager and Filev 1994). The consequent part of the fuzzy rules in the so-called Takagi–Sugeno (TS) form is represented by mathematical functions (usually linear). The structure of the TS fuzzy system can also be represented as a neural network with a specific (five layer) composition (Fig. 1). Therefore, these systems are also called neuro-fuzzy (NF).

The number of fuzzy rules and inputs (which in case of classification problems are also called features or attributes) is also a part of the structure.

The first layer consists of neurons corresponding to the membership functions of a particular fuzzy set. This layer takes the inputs, *x* and gives as output the degree, *μ* to which these fuzzy descriptors are satisfied. The



Evolving Fuzzy Systems, Fig. 1 Structure of the (neuro-fuzzy) system of TS type

second layer represents the antecedent parts of the fuzzy rules. It takes as inputs the membership function values and gives as output the firing level of the i th rule, τ_i . The third layer of the network takes as inputs the firing levels of the respective rule, τ_i and gives as output the normalized firing level, λ_i as “center of gravity” (Klir and Folger 1988) of τ_i . As an alternative one can use the “winner takes all” operator. This operator is used usually in classification, while the “center of gravity” is preferred for time-series prediction and general system modeling and control. The fourth layer aggregates the antecedent and the consequent part that represents the local sub-systems (singletons or hyper planes). Finally, the last fifth layer forms the total output of the NF system. It performs a weighed summation of local sub-systems.

Fuzzy system parameters Parameters of the NF system of TS type include the center, c and spread, σ of the Gaussians or parameters of the triangular (or trapezoidal) membership functions. An example of a Gaussian type membership function can be given as:

$$\mu = e^{-\frac{1}{2}(\frac{d}{\sigma})^2} \quad (2)$$

where d denotes distance between a data sample (point in the data space) and a proto-type/cluster center (focal point of a fuzzy set);

r is the radius of the cluster (spread of the membership function).

Note that the distance can be represented by Euclidean (the most typical example), Mahalanobis (Hastie et al. 2001), cosine etc. forms.

These parameters are associated with the antecedent part of the system. Consequent part parameters are coefficients of the (usually) linear functions, singleton coefficients or coefficients of more complex functions (e. g. exponential) if such ones are used.

$$y_i = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n \quad (3)$$

where a denotes parameters of the consequent part; x denote the inputs (features); i is the index of the i th fuzzy rule; n is the number (dimensionality) of the inputs (features).

Potential Potential is a mathematical measure of the data density. It is calculated at a data point, z and represents numerically the accumulated proximity (*density*) of the data surrounding this data point. It resembles the probability distribution used in so-called Parzen windows (Hastie et al. 2001) and is described in (Yager and Filev 1993; Chiu 1994) by a Gaussian-like function:

$$P(z) = e^{-\frac{1}{2\sigma^2}z^2} \quad (4)$$

where $z = [x, y]$ denotes the joint (input/output) vector;

$\bar{\sigma}_k^2 = \frac{1}{k-1} \sum_{i=1}^{k-1} d^2(z_k, z_i)$ is the variance of the data in terms of the cluster center.

In (Angelov and Filev 2004; Angelov 2004a) the Cauchy function is used which has the same properties as the Gaussian but is suitable for recursive calculations.

$$P(z) = \frac{1}{1+\sigma^2}. \quad (5)$$

Age of a cluster or fuzzy rule The age of the (evolving) cluster is defined as the accumulated time of appearance of the samples that form the cluster which support that fuzzy rule.

$$A^i = k - \frac{\sum_{l=1}^{S_k^i} k_l}{S_k^i} \quad (6)$$

where k denotes the current time instant; S_k^i denotes the support of the cluster that is the number of data samples (points) that are in the zone of influence of the cluster (formed by its radius). It is derived by simple counting of data samples (points) at the moment of their arrival (when they are first read) and assigned to the nearest cluster (Angelov and Filev 2005).

The values of A vary from 0 to k and the derivative of A in respect to time is always less or equal to 1 (Angelov et al. 2007b). An “old” cluster (fuzzy rule) has not been updated recently. A “young” cluster (fuzzy rule) is one that has predominantly new samples or recent ones. The (first and second) derivatives of the *age* are very informative and useful for detection of data “*shift*” and “*drift*” (Angelov et al. 2007b).

Definition of the Subject

Evolving Fuzzy Systems (EFS) are a class of Fuzzy Rule-based (FRB) and Neuro-Fuzzy (NF) systems that have both their parameters and underlying structure self-adapting, self-developing, self-learning from the data in on-line mode and, possibly, in real-time. The concept was conceived at the beginning of this century (Angelov and Buswell 2001; Angelov 2002). Parallel investigations have led to similar

developments in neural networks (NN) (Kasabov 2001; Kasabov and Song 2002). EFS have the significant advantage compared to the evolving NN of being linguistically tractable and transparent. EFS have been instrumental in the emergence of new branches of *evolving* clustering algorithms (Angelov 2004a), *evolving* classifiers (Lughofer et al. 2007; Angelov et al. 2007a), *evolving* time-series predictors (Angelov and Filev 2004; Leng et al. 2005), *evolving* fuzzy controllers (Angelov 2004b), *evolving* fault detectors (Filev and Tseng 2006) etc. Over the last years EFS has demonstrated a wide range of applications spanning robotics (Zhou and Angelov 2007) and defense (Carline et al. 2005) to biomedical (Xydeas et al. 2006) and industrial process (Filev et al. 2000) data processing in real-time, new generations of self-calibrating, self-adapting sensors (Macias et al. 2006; Macias-Hernandez et al. 2007), speech (Jones and Xydeas 2006) and image processing (Memon et al. 2006) etc. EFS have the potential to revolutionize such areas as autonomous systems (Valavanis 2006), intelligent sensors (Kordon 2006), early cancer detection and diagnosis; they are instrumental in raising the so-called machine intelligence quotient (Zadeh 1993) by developing systems that self-adapt in real-time to the dynamically changing environment and to internal changes that occur in the system itself (e. g. wearing, contamination, performance degradation, faults etc.). Although the terms *intelligent* and *artificial intelligence* have been used often during the last several decades the technical systems that claim to have such features are in reality far from true intelligence. One of the main reasons is that true intelligence is evolving, it is not fixed. EFS are the first mathematical constructs that combine the approximate reasoning typical for humans represented by the fuzzy inference with the dynamically evolving structure and respective formal mathematically sound learning mechanisms to implement it.

Introduction

Fuzzy Sets and Fuzzy Logic were introduced by Lotfi Zadeh in 1965 in his seminal paper (Yager 2006). During the last decade of the previous

century there was an increase of the various applications of fuzzylogic-based systems mainly due to the introduction of fuzzy logic controllers (FLC) by Ebrahim Mamdani in 1975 (Mamdani and Assilian 1975), the introduction of the fuzzily blended linear systems construct called Takagi–Sugeno (TS) fuzzy systems in 1985 (Takagi and Sugeno 1985), and theoretical proof that FRB systems are universal approximators (that is any arbitrary non-linear function in the $[0; 1]$ range can be asymptotically approximated by a FRB system (Wang 1992)). Historically, the FRB systems were first being designed based entirely or predominantly on human expert knowledge (Mamdani and Assilian 1975; Yager 2006). This offers advantages and was a novel technique at that time for incorporating uncertain, subjective information, preferences, experience, intuition, which are difficult or impossible to be described otherwise. However, it poses enormous difficulties for the process of designing and routine use of these systems, especially in real industrial environments and in on-line and real-time modes. TS fuzzy systems made possible the development of efficient algorithms for their design not only in design not only in off-line, but also in on-line mode (Angelov et al. 2004a). This is facilitated by their dual nature – they combine a fuzzy linguistic premise (antecedent) part with a functional (usually linear) consequent part (Takagi and Sugeno 1985). With the invention of the concept of EFS (Angelov and Buswell 2001; Angelov 2002) the problem of the design was completely automated and data-driven. This means, EFS systems self-develop their model, respectively system structure as well as adapt their parameters “from scratch” on the fly using experimental data and efficient recursive learning mechanisms. Human expert knowledge is not compulsory, not limiting, not essential (especially if it is difficult to obtain in real-time). This does not necessarily mean that such knowledge is prohibited or not possible to be used. On the contrary, the concept of EFS makes possible the use of such knowledge in initialization stages, even during the learning process itself, but this is not essential, it is optional. Examples of EFS are intelligent sensors for oil refineries (Macias et al. 2006; Macias-

Hernandez et al. 2007), autonomous self-localization algorithms used by mobile robots (Zhou and Angelov 2006; Zhou and Angelov 2007), smart agents for machine health monitoring and prognosis in the car industry (Filev and Tseng 2006), smart systems for automatic classification of images in CD production process (Lughofer et al. 2007) etc. This is a new promising area of research and new applications in different branches of industry are emerging.

Evolving Clustering

Data Clustering and, Fuzzy Clustering in particular, are methods for grouping the data based on their similarity, density in the data space and proximity. Partitioning of the data into clusters can be done off-line (using a batch set of data, performing iterative computations over this set, minimizing certain criteria/cost function) or on-line, incrementally. Examples of incremental clustering approaches are self-organizing maps (SOM) conceived by Teuvo Kohonen in the early 1980s (Kohonen 1995), adaptive resonance theory (ART) by Stephen Grossberg conceived in the same period (Carpenter and Grossberg 2003) etc. Clustering is a type of unsupervised learning technique where the correct examples are not provided. Usually, the number of clusters is pre-specified, e. g. in SOM the number of nodes of the map is pre-defined; the number of neighbors, k in the k -nearest neighbor clustering method (Hastie et al. 2001) is also supposed to be provided, the number C in the fuzzy c -means (FCM) fuzzy clustering algorithm by Jim Bezdek should also be provided (Bezdek 1974). Usually these approaches rely on a threshold and are very sensitive to the specific values of this threshold. Most of the existing approaches are also mean-based (i. e. they use the mean of all data or mean of groups of data). The problem is that the mean is a virtual (non-existing and possibly infeasible) point in the data space.

In contrast, the evolving clustering method eClustering conceived in the last decade (Angelov 2004a) does not need the number of clusters, the threshold or any other parameter to

be pre-specified. It is parameter-free and starts “from scratch” to cluster the data based on their density distribution alone. It is based on the recursive calculation of the potential (5). eClustering is prototype-based (some of the data points are used as prototypes of cluster centers). The procedure of the evolving clustering approach starts from scratch assuming that the first available data point is a center of a cluster. This assumption is temporary and if a priori knowledge exists the procedure can start with an initial set of cluster centers that will be further refined. The coordinates of the first cluster center are formed from the coordinates of the first data point ($z_1^* \leftarrow z_1$). Its potential is set to the ideal value, $P_1(z_1) \rightarrow 1$. Starting from the **next** data point which is read in real-time, the following steps are performed for each new data point:

- calculate its potential, $P_k(z_k)$;
- update the potential of the existing cluster centers (because their potential has been affected by adding a new data point);
- compare the potential of the new data point with the potential of the previously existing centers. On the basis of this comparison and the membership of the existing clusters one of the following actions is taken:

(**add** a new cluster center based on the new data point) **OR** (**remove** the cluster that describes well the new point which brings an increment to the potential) **AND** (**replace** it with a cluster formed around the new point) **OR** (**ignore** (do not change the cluster structure)).

The process is illustrated in Fig. 2 for the data of NOx emissions from a car exhaust (Angelov et al. 2005).

One can see that the clustering evolves (number of clusters increases from two to three, their position and their radius has changed. Note that in this experiment only two normalized to the range [0; 1] inputs (features), namely the engine output torque in N/m, x_1 and pressure in the second cylinder in Pa, x_2 are used. For more details on eClustering, please, consult the papers from the bibliography, and especially (Angelov 2002;

Angelov and Filev 2004; Angelov 2004a; Angelov et al. 2007a).

Evolving TS Fuzzy Systems

TS fuzzy systems [as illustrated in Fig. 1 and described in a very general form in Eq. (1)] were first introduced in 1985 (Takagi and Sugeno 1985) in the form:

$$\mathfrak{R}_i: \text{IF } (x_1 \text{ is } A_{i1}) \text{ AND } \dots \text{ AND } (x_n \text{ is } A_{in}) \\ \text{THEN } (y_i = a_{i0} + a_{i1}x_1 + \dots + a_{in}x_n) \quad (7)$$

where A_i denotes the i th fuzzy rule ($i = [1, R]$); R is the number of fuzzy rules; x is the input vector; $x = [x_1, x_2, \dots, x_n]^T$; A_{ij} denotes the antecedent fuzzy sets, $j \in \{1, n\}$; y_i is the output of the i th linear subsystem; a_{il} are its parameters, $l \in \{0, n\}$.

The structure of the TS system (number of fuzzy rules), antecedent part of the rules, number of inputs etc., are supposed to be known and are fixed. The data may be provided to the TS system in off-line or in on-line manner. Different data-driven techniques were developed to identify the best in terms of certain (local or global) error minimization criteria such as using a (recursive) least squares technique (Takagi and Sugeno 1985), using genetic algorithms (Setnes and Roubos 2000; Angelov and Buswell 2003) etc. The overall output is found to be a weighted sum of local outputs produced by each fuzzy rule:

$$y = \sum_{i=1}^R \lambda^i y^i. \quad (8)$$

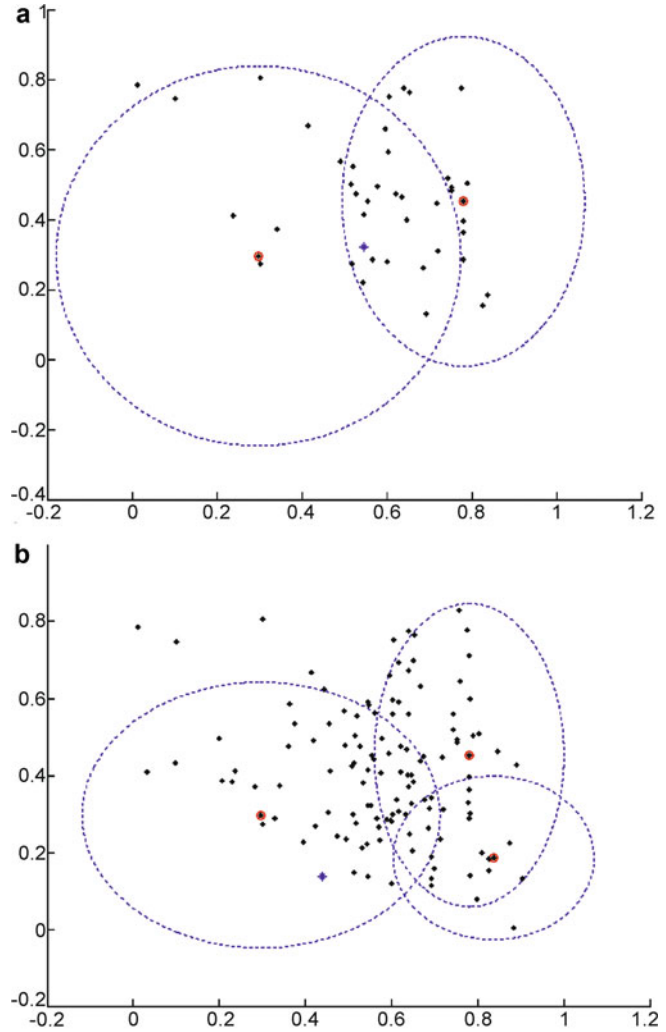
Where the weights, λ represent the normalized firing level of the respective fuzzy rule and can be determined by:

$$\lambda^i = \frac{\sum_{j=1}^n T \mu_j^i(x_j)}{\sum_{l=1}^R \sum_{j=1}^n T \mu_j^l(x_j)}. \quad (9)$$

In a vector form the above equations can be represented as:

Evolving Fuzzy Systems,

Fig. 2 The Evolving Clustering method applied to data concerning NOx emissions; **(a)** top plot—after 43 samples are read (after 43 s because the sampling rate is 1 sample/second or 1 Hz); **(b)** bottom plot—after 124 samples are read (after 124 s)



$$y = \psi^T \theta \quad (10)$$

where $\psi = [\lambda_1 x_e^T, \lambda_2 x_e^T, \dots, \lambda_R x_e^T]^T$ is the vector of weighted extended inputs; $x_e^T = [1, x^T]$;

$\theta = [\pi_1^T, \pi_2^T, \dots, \pi_R^T]^T$ is the vector of parameters;

$$\pi^i = \begin{bmatrix} \alpha_{01}^i & \alpha_{02}^i & \dots & \alpha_{0m}^i \\ \alpha_{11}^i & \alpha_{12}^i & \dots & \alpha_{1m}^i \\ \dots & \dots & \dots & \dots \\ \alpha_{n1}^i & \alpha_{n2}^i & \dots & \alpha_{nm}^i \end{bmatrix}$$

are the parameters of the m local linear sub-systems.

The assumption that the TS fuzzy system structure has to be known a priori was for the first time questioned in (Angelov and Buswell 2001; Angelov 2002) and ultimately in (Angelov and Filev 2004) with the proposal of evolving TS (eTS) systems. In (Angelov et al. 2004b) a further extension of the eTS system was proposed, namely that they can have many outputs. In this way, the multi-input-multi-output eTS systems were introduced (MIMO-eTS).

eTS is a very flexible and powerful tool for time-series prediction, prognosis, modeling non-stationary phenomena, intelligent sensors etc. The algorithm for its learning from streaming data in real-time has two basic phases, which can both be performed very quickly (in onetime step between

the arrival of two data samples – the current one and the next one). The learning mechanism proposed in (Angelov and Filev 2004) is computationally very efficient because it is fully recursive. The two phases include:

- (a) Data space partitioning and based on this forming and update of the fuzzy rule-base structure;
- (b) Learning parameters of the consequent part of the fuzzy rules.

Note that the partitioning of the data space serves in eTS identification a different purpose compared to the purpose of data space partitioning in eClustering. In eTS there are outputs and the aim is to find such (perhaps overlapping) clustering of the input-output joint data space that fragments the input-output relationship into locally valid simpler (possibly linear) dependences. In eClustering the aim is to cluster the input data space into distinctive regions. Other than that, the first phase of the eTS model identification is the same as the procedure in the eClustering method described above.

The second phase of the learning is parameter identification. It can be performed using a fuzzily weighted version (Angelov and Filev 2004) of the well-known recursive least squares (RLS) method (Ljung 1987). One can perform either local (11) or global (12) identification by minimizing different cost functions (Angelov and Filev 2004):

$$J_L = \sum_{i=1}^R (Y - X^T \pi_i)^T A_i (Y - X^T \pi_i) \quad (11)$$

$$J_G = (Y - \Psi^T \theta)^T (Y - \Psi^T \theta). \quad (12)$$

In one of the cases (when a local cost function is used) the result will be a better approximation locally of the overall non-linear function by the local linear sub-models. The pay-off is, however, a poorer overall approximation. This is, however, compensated by a simpler and computationally more efficient procedure (if we use a locally valid cost function the covariance

matrices of much smaller size can be used and they require much less memory space and time to perform computations) (Angelov and Filev 2004).

Evolving Fuzzy Classifiers

Classification is a problem that has been well studied and a large number of conventional approaches exist to address this problem (Hastie et al. 2001). Most of them, however, are designed to operate in batch mode and do not change their structure on-line (do not capture new patterns that may be present in the streaming data once the classifier is built). Off-line pre-trained classifiers may be good for certain scenarios, but they need to be redesigned or retrained if the circumstances change. There are also so-called incremental (or on-line) classifiers which work on a “sample-by-sample” basis and only require the features of that sample plus a small amount of aggregated information (a rule-base, a small number of variables needed for recursive calculations). They do not require all the history of the data stream (all previously seen data samples). Sometimes they are also called one-pass (each sample is processed only once at a time and is then discarded from the memory).

FRB systems have been successfully applied to a range of classification tasks including, but not limited to, decision making, fault detection, pattern recognition, image processing (Kuncheva 2000). FRB systems have become one of the alternative frameworks for classifier design together with the more established Bayesian classifiers, decision trees (Hastie et al. 2001), neural network-based classifiers (Nauck and Kruse 1997), and support-vector machines (SVM) (Vapnik 1998). The task of the classifier is to map the set of features of the sample data onto the set of class labels. A particular advantage of the FRB classifiers is that they are linguistic in form while being also proven universal approximators (Wang 1992).

In the framework of the concept of *evolving fuzzy systems* a family of evolving fuzzy

classifiers, eClass was proposed in (Lughofer et al. 2007; Angelov et al. 2007a, b). The first type of evolving fuzzy classifiers, eClass0 has the typical structure of a fuzzy classifier (Kuncheva 2000) that differs from structure (1) by the consequent part only:

$$\begin{aligned} R^i: & \text{IF } (Feature_1 \text{ is close to } prototype_1^i) \\ & \text{AND} \dots \text{AND } (Feature_n \text{ is close to } prototype_n^i) \\ & \text{THEN } (ClassLabel^i) \end{aligned} \quad (13)$$

The output of eClass0, in the same way as typical fuzzy classifiers (Kuncheva 2000) provides the label of the class (0, 1 etc.) directly. In this sense, it is not a TS fuzzy system, but is closer to the Mamdani-type fuzzy systems (Mamdani and Assilian 1975). The main difference of the eClass0 from the typical classifiers (Kuncheva 2000) is its ability to evolve, to expand the set of fuzzy rules that enables it to capture new data patterns, to adapt to possibly changing characteristics of the streaming data (Angelov et al. 2007a, b). The inference in eClass0 is produced using the so-called “winner takes all” rule (Kuncheva 2000; Hastie et al. 2001).

$$\begin{aligned} Label &= Label^{i*}; i^* \\ &= \arg\max_{i=1}^R \left(\sum_{j=1}^n T \mu_j^i(x_j) \right). \end{aligned} \quad (14)$$

It is much easier and faster to build and train eClass0 in real-time, but the results of the classification can be further significantly improved if the classifier structure is assumed to be of TS-type. eClass is designed for on-line applications with an evolving (self-developing) FRB structure. The antecedents of the FRB are formed from the data stream around highly descriptive focal points (prototypes) per class in the input-output space. The features vector, $\mathbf{x}$ is augmented with the class label, L to form a joint input-output vector $\mathbf{z} = [\mathbf{x}^T, L]^T$. The eClustering algorithm is applied *per class* (not over all the data). In this way, information granules (primitive forms of knowledge) (Ishibuchi et al. 2004) are formed **in real-time** around descriptive data samples, represented

linguistically by fuzzy sets. This on-line algorithm works similarly to adaptive control (Astroem and Wittenmark 1994) and estimation (Kalman 1960) – in the period between two samples two phases are performed: (1) class prediction (classification); (2) classifier update or evolution. During the first phase the class label is not known and is being predicted; during the second phase, however, it is known and is used as supervisory information to update the classifier (including its **structure evolution** as well as its parameters update).

An alternative structure of the fuzzy classifier, eClass1 is based on the TS-type fuzzy system which has a consequent part of functional type as described in (7). The architecture of eClass1 differs significantly from the architecture of eClass0 and the typical FRB (Kuncheva 2000). It performs a regression over the features. Having in mind that the classification surface in a data stream changes dynamically the goal of the evolving fuzzy classifier eClass1 is to evolve a rule-base which takes these changes into account by adapting parameters of the FRB (spreads, consequent parameters) as well as the focal points and the size of the rule-base. The output of each rule is a real (not integer as in the typical fuzzy classifiers) value, which if normalized represents the possibility of a data sample to be of certain class (Angelov et al. 2007a, b):

$$\bar{y}^i = \frac{y^i}{\sum_{i=1}^R y^i}. \quad (15)$$

The overall output of the classifier is then taken as a weighted average (not as winner takes all as in typical fuzzy classifiers) of the normalized outputs of each fuzzy rule:

$$y = \sum_{i=1}^R \frac{\sum_{j=1}^n T \mu_j^i}{\sum_{i=1}^R \sum_{j=1}^n T \mu_j^i} \bar{y}^i. \quad (16)$$

This output is then used to discriminate between the classes. If the problem has two classes (A and B) then the target values are, obviously, 0 for one of the two classes (e. g. Class A) and

1 for the other one (Class B) or vice versa. To discriminate in this case one can simply use a threshold of 0.5. All the outputs that are above 0.5 are being classified as Class B while all the outputs below 0.5 are classified as Class A or vice versa:

$$\begin{aligned} &\text{IF } (y > 0.5) \\ &\text{THEN } (\text{Class } A) \\ &\text{ELSE } (\text{Class } B) \end{aligned} \quad (17)$$

When the problem has more than two classes, one can apply MIMO eTS where each of the K outputs corresponds to the possibility that a data sample belongs to a certain class (as discussed above). It is interesting to note that it is possible to use MIMO eTS for a two-class problem. In order to do this, one needs to have vectors that represent the target outputs, for example $\mathbf{y} = [1 \ 0]$ for Class A and $\mathbf{y} = [0 \ 1]$ for Class B or vice versa. In eClass1-MIMO the label is determined by the highest value of the discriminator, $\bar{y}_l$:

$$\text{Label} = \text{Label}^{i^*}; i^* = \arg \max_{l=1}^K \bar{y}_l \quad (18)$$

where K denotes the number of classes.

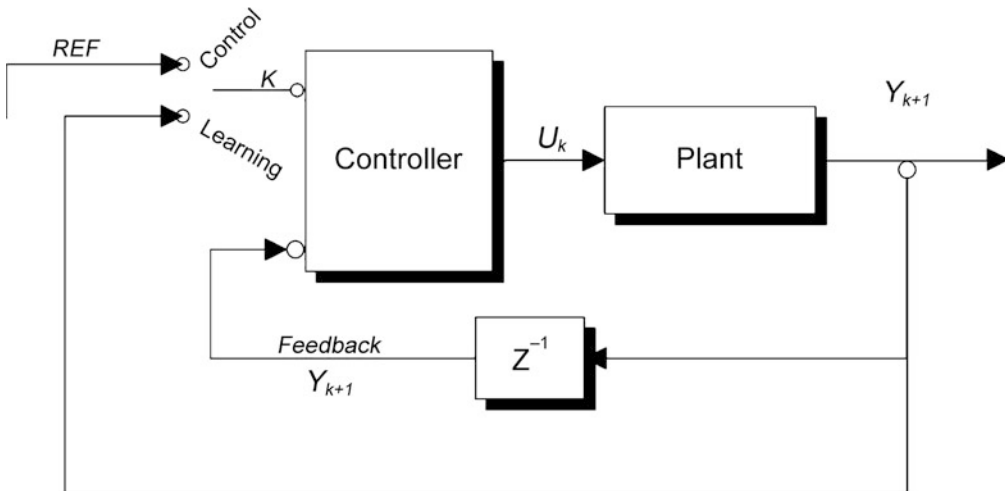
Evolving Fuzzy Controllers

Fuzzy logic controllers have been applied in a range of industrial processes (Procyk and Mamdani 1979; Lin et al. 2001) around the world including in home appliances (Badami and Chbat 1998). The structure of the controller, however, is often decided in an ad hoc manner (Mamdani and Assilian 1975; Procyk and Mamdani 1979) and parameters are tuned off-line using various numerical techniques such as genetic algorithms for example (Goldberg 1989; Shimojima et al. 1995; Cordon et al. 2004). In reality, however, even if a validation test has been made beforehand, there is no guarantee that a controller designed in this way will perform satisfactory if the object of control or its environment change (Angelov 2004b). The reasons could be aging,

wearing, change in the mode of operation or development of a fault, seasonal changes in the environment etc. An effective mechanism for tackling such problems known from the classical control theory is adaptation (Astroem and Wittenmark 1994). It is well developed for the linear models and controllers (Kailath et al. 2000), but not for a general (very often highly non-linear, complex and uncertain) case (Angelov 2002). Adaptive control theory assumes a linear model with a fixed structure and applies to parameters only (Astroem and Wittenmark 1994; Kailath et al. 2000).

The concept of evolving fuzzy systems has been applied to the control problem in (Angelov 2002, 2004b) in terms of self-developing the structure of a controller from experimental data in a data-driven manner based on the indirect adaptive learning scheme proposed initially by Psaltis (Psaltis et al. 1988) and developed further using NN by Anderson (Andersen et al. 1994). The indirect learning (IL) control scheme is based on the approximation of the inverse dynamics of the plant. The IL-based control scheme is a model-free concept. It feeds back the integrated (or delayed one-step back) output signal instead of feeding back the error between the plant output and the reference signal as represented in Fig. 3.

Figure 3 represents only the basic concept of the approach. It has two phases and the switching between them can be represented by an imaginary switch knob. When the imaginary knob, K is in position “1” the controller is used and we are in phase “Control”. When the imaginary knob is in position “2” the controller learns, self-develops and we are in phase “Learning”. During the supervisory learning phase, the true output signal ($\mathbf{y}_{k+1}$) at the time-instant($k + 1$) is fed back and the knob is in position “2”. The controller also receives a signal that is a delayed true output, $\mathbf{y}_k$. The controller has as an output the value of the control signal, $\mathbf{u}_k$. During the control phase (when the knob is in position “1”) the input is determined entirely based on the reference signal (ref) as an alternative to the predicted next step output, $\mathbf{y}_{k+1}$. In this way, the controller already trained in the previous learning phases produces such a control



Evolving Fuzzy Systems, Fig. 3 Indirect learning-based control scheme

signal (u_k), which brings the output of the plant at the next time step (y_{k+1}) close to the reference signal (ref).

The IL scheme was taken further in (Angelov 2002; Angelov 2004b) by implementing the controller as an evolving FRB system of TS-type. The original works realized the controller as a NN that was trained *off-line* based on a batch set of training data for the control action and output triplets of the form $[y_k, y_{k+1}, u_k]$ for $k = 1, 2, \dots, N$; where N denotes the number of training data samples. However, learning techniques for NN are iterative and, therefore training of the NN-controller as described in (Psaltis et al. 1988; Andersen et al. 1994) is performed *off-line*. Additionally, NN suffers from the important disadvantage in comparison to the FRB systems that they are not transparent. In (Angelov 2002, 2004b) the basic scheme of IL control is taken further by adding a disturbance and by using eTS to realize the controller. This scheme was implemented on temperature-control problems.

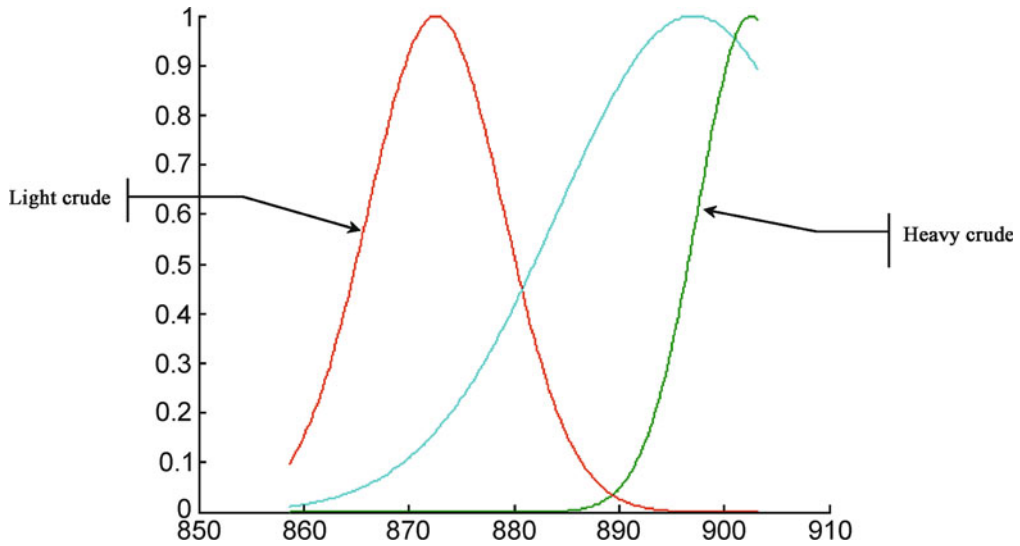
Application Case Studies

Self-Calibrating Intelligent Sensors for Process Industries

So-called intelligent or inferential sensors have been adopted by the process industries

(chemical, petro-chemical, manufacturing) for several decades (Fortuna et al. 2007). The main reason is that they provide accurate real time estimates of difficult to measure otherwise parameters or can replace (substitute) expensive measurements such as gas emissions, biomass, melt index, etc. They use as inputs the available (“hard”) sensors for easy to measure physical variables, such as temperatures, pressures, and flows which are also cheaper. The main disadvantage of the currently existing inferential or “soft” sensors is that significant efforts must be made based on batch sets of data to develop and maintain the mathematical models that support them (neural networks, statistical models etc.). Any process changes outside the conditions used for off-line model development can lead to significant performance deterioration which, in turn, requires maintenance and re-calibration.

Evolving Fuzzy Systems offer an effective opportunity to develop “soft” sensors that are more flexible, self-calibrating, and thus, more “intelligent” (Kordon 2006). Several applications of EFS-based soft sensors, in particular for oil refineries (Macias et al. 2006; Macias-Hernandez et al. 2007) and propylene production (Angelov et al. 2008) were reported. An important advantage of the evolving sensors is that they extract human-interpretable knowledge in the form of linguistic fuzzy rules. For example, Fig. 4



Evolving Fuzzy Systems, Fig. 4 Fuzzy sets for different fractions of the crude that contribute to the different quality of the end product (naphtha in this case) can be extracted automatically in real-time from the data stream using eSensor

illustrates membership functions of the fuzzy sets (in values in the range $[0;1]$ on the vertical axis) that describe the density of the crude, d in gram per liter (g/l) on the horizontal axis. The evolving fuzzy sensor (eSensor) implemented in the oil refinery at Santa Cruz, Tenerife, Spain predicts in real-time the temperature of the heavy naphtha (hn) T^{hn} , °C in degrees Celsius when it evaporates 95% liquid volume, according to the ASTM D86-04b standard based on real-time measurements of:

- The pressure of the tower, p , measured in $\text{kg}/\text{cm}^2\text{g}$.
- The amount of the product taking off, P , represented in %.
- The density of the crude, d in g/l (illustrated in Fig. 4).
- Temperature of the column overhead, T^{co} in °C.
- Temperature of the naphtha extraction, T^{ne} in °C.

An expert in the area of oil refining processes can easily visually distinguish between the heavy crude and light crude represented by respective membership functions derived automatically from the data in real-time.

Adaptive Real-Time Classifiers (Image, Land-Mark, Robotic)

Presently the data that are to be processed in industry, in defense and other real-life applications are not only huge in volume, but very often they are in the form of data *streams* (Domingos and Hulten 2001). This requires not only precise classifiers, but also dynamically evolvable classifiers. For example, in mobile robotics, an autonomous vehicle produces a video stream while operating in a completely unknown environment that needs to be processed (Zhou and Angelov 2006; Zhou and Angelov 2007). The Evolving Clustering method was used to automatically generate a fuzzy rule-base that describes the landmarks discovered without any prior learning, “on the fly” by a mobile robot Pioneer 3DX (Pioneer-3DX 2004) exploring a completely unknown environment (Zhou and Angelov 2007). The mobile robot is using its on-board pan-tilt zoom camera to produce a video stream. The frames were grasped and processed by the eClustering algorithm based on a 3-dimensional color-vector (R, G, B). Fuzzy rules of the following form were extracted from the data automatically:

Note that the landmarks that were identified automatically represented real objects on the

route of the mobile robot such as the underpass of Lancaster University from the example. They were identified to be distinctive from the surrounding background by eClustering. The fuzzy rule base (Fig. 5) has evolved “from scratch” based on the video information and the data distribution only.

Predictive Models (Air-Conditioning, Financial Time-Series, Benchmark Data Sets)

There are different techniques that can be used for predictive models, such as ARMAX models (Ljung 1987), neural networks (Hornby 1974) non-evolving (fixed structure) fuzzy rule-based models (Takagi and Sugeno 1985; Yager and Filev 1994). Evolving fuzzy systems, however, offer additionally the capability to have a predictor that evolves following the dynamic changes of the data by gradually adapting not only its parameters, but also its structure. In (Angelov and Buswell 2002) the problem of predicting the characteristic temperature difference across a coil in a heat exchanger of an air-conditioning unit installed in a real building in Iowa, USA is considered. The evolving fuzzy rule-based model develops its structure and parameters based on the data of the flow rate entering the coil, moisture content of the air entering the coil, temperature of the chilled water, and control signal to the valve as illustrated in Fig. 6. The model proved to work satisfactorily in all season conditions due to its ability to adapt to the changes in the environment (different seasons) as demonstrated in Fig. 6c.

When pre-trained (based on 400 samples) and fixed as both structure and parameters the performance deteriorated unacceptably in changing seasonal conditions as seen in Fig. 6a. A partial re-training improved significantly the results (Fig. 6b), but still this was less valid when the season changed (at around sample 912) and the model structure evolution has stopped (at sample 1000). When, the model structure evolution continued uninterrupted the result was a satisfactory performance in all seasons as seen in Fig. 6c.

Fault Detection and Prognostics

Evolving clustering and eTS fuzzy systems were applied in Ford Motor Company to machine health monitoring and prognosis in (Filev and Tseng 2006). The ability of the evolving clustering method to form new clusters that represent different operating modes of a machine was exploited and different types of faults (incipient or drastic) were automatically identified based on the difference in the cluster formation. A prediction of the direction of movement of the cluster centers was used for prediction of possible faults and of the end-of-life of the machine.

Speech Signal Reconstruction

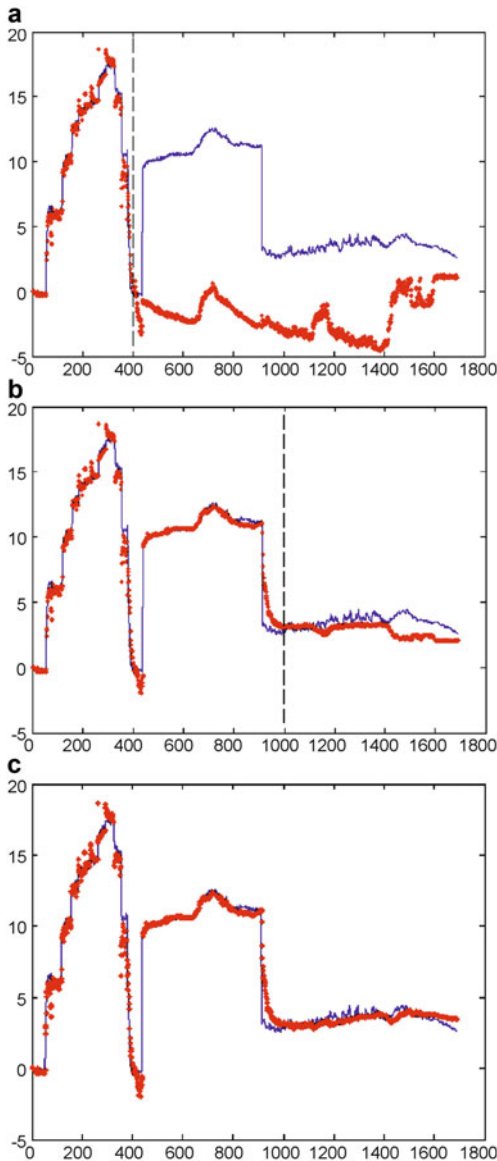
The ETS fuzzy system is used in (Jones et al. 2006) for error concealment in the next-generation Voice over Internet Protocol (VoIP) communication receivers. It is used in combination with parametric speech coders of analysis-by-synthesis-type. eTS MIMO (Angelov et al.

Rule_i: IF (μ^R_{il} is close to 123) AND... AND(μ^B_{in} is close to 84)

THEN the Frame_j is like the Landmark_i:



Evolving Fuzzy Systems, Fig. 5 Fuzzy rule describing a Landmark (underpass at Lancaster University campus) that was discovered automatically by eClustering using video streaming data



Evolving Fuzzy Systems, Fig. 6 A predictive model of the characteristic temperature difference across a coil in a heat exchanger of an air-conditioning unit installed in a real building in Iowa, USA. (a) An off-line pre-trained model used in two different seasons (summer and spring); (b) Evolving FRB model trained and used during the summer (up to sample 1000) and then having its structure fixed during the spring season; (c) Evolving FRB model left to evolve (self-develop) during the whole period of usage (both spring and summer seasons).

2004b) is used to predict the missing values of the linear spectral pairs (LSP) that will allow one to reconstruct the lost in transmission packets. The eTS fuzzy model used ten inputs (current LSP

parameter values) and ten outputs (predicted one step/20 milliseconds ahead LSP values). This research was a joint work between Lancaster University and Nokia-UK and aims the development of the next generation intelligent decoders at the receiver that will be able to conceal lost packets with a size of 80–160 ms without significant deterioration to the quality of service (QoS) in VoIP transmission.

Future Directions

The area of evolving fuzzy systems is in its infancy and has already demonstrated a remarkable success in addressing some of the most vibrating issues of the development, application and implementation of truly intelligent systems in a wide variety of branches of industry and real-life problems (Angelov et al. 2006). It opens the door for future developments that are related to the areas of autonomous systems, early cancer diagnosis and prognosis of the progression, even to the identification of structural changes in biological cells that correspond to the evolution of the disease. In the area of intelligent self-maintaining sensors the process industry can benefit from more flexible and smarter solutions. The problems that are yet to be addressed and can mark the future development of this vibrant area are; (1) collaboration aspects between two or more evolving fuzzy system-based intelligent systems (autonomous robots, intelligent sensors etc.); (2) further flexibility of the systems in terms of real-time self-analysis, optimal features and input selection, rule aggregation mechanism adaptation etc.; (3) even more flexible system structure architectures such as hierarchical, decentralized; (4) more robust learning algorithms that take care of missing data, different sampling intervals etc.

From a broader prospective, the future developments of this discipline will influence and are closely related to similar developments in the area of communication networks (self-adaptive networks (Sifalakis and Hutchison 2004)), self-validating soft sensors (Qin et al. 1997), autonomous aerial, ground-based, and underwater vehicles (Azimi-Sadjadi et al. 2002; Kanakakis et al.

2004; Liu and Meng 2004) etc. The area is closely related to the developments in the area of neural networks (Hornby 1974; Lin et al. 2001), so-called autonomous mental development (Bonarini et al. 2006) and cognitive psychology (Massaro 1991), mining data streams (Domingos and Hulten 2001). One can also expect more hardware implementations (the first hardware implementation of eClustering was reported in 2005 (Angelov and Everett 2005)). From the point of view of mathematical fundamentals and learning it is also closely related to adaptive filters theory (Widrow and Stearns 1985) and the recent developments in particle filters (Angelov et al. 2007b) will certainly influence the future, more efficient techniques that will be developed in this emerging and highly potential branch of research.

Acknowledgments The author would like to thank Mr. Xiaowei Zhou for his assistance in producing the illustrative material, Dr. Jose Macias Hernandez for kindly providing real data from the oil refinery CEPISA, Santa Cruz, Tenerife, Spain, Dr. Richard Buswell, Loughborough University and ASHRAE(RP-1020) for the real air conditioning data, and Dr. Edwin Lughofer from Johannes Kepler University of Linz, Austria for providing real data from car engines.

Bibliography

Primary Literature

- Andersen HC, Teng FC, Tsoi AC (1994) Single net indirect learning architecture. *IEEE Trans Neural Netw* 5: 1003–1005
- Angelov P (2002) *Evolving rule-based models: a tool for Design of Flexible Adaptive Systems*. Springer, Heidelberg
- Angelov P (2004a) An approach for fuzzy rule-base adaptation using on-line clustering. *Int J Approx Reason* 35(3):275–289
- Angelov PP (2004b) A fuzzy controller with evolving structure. *Inf Sci* 161:21–35
- Angelov P, Buswell R (2001) Evolving rule-based models: a tool for intelligent adaptation. In: *Proceedings of the joint 9th IFSA world congress and 20th NAFIPS intern conference*, Vancouver, 25–28 July 2001. IEEE Press, USA, pp 1062–1066
- Angelov P, Buswell R (2002) Identification of evolving rule-based models. *IEEE Trans Fuzzy Syst* 10(5): 667–677
- Angelov P, Buswell R (2003) Automatic generation of fuzzy rule-based models from data by genetic algorithms. *Inf Sci* 150(1/2):17–31
- Angelov P, Everett M (2005) *EvoMap: on-Chip implementation of intelligent information modelling using EVOLving MAPping*. Lancaster University, Lancaster, pp 1–15
- Angelov P, Filev D (2004) An approach to on-line identification of evolving Takagi–Sugeno models. *IEEE Trans Syst Man Cybern part B Cybern* 34(1):484–498
- Angelov P, Filev D (2005) *Simpl_eTS: a simplified method for learning evolving Takagi–Sugeno fuzzy models*. In: *Proceedings of the 2005 IEEE international conference on fuzzy systems FUZZ-IEEE – 2005*, vol 2005. Reno, pp 1068–1073
- Angelov P, Victor J, Dourado A, Filev D (2004a) On-line evolution of Takagi–Sugeno fuzzy models. In: *Proceedings of the 2nd IFAC workshop on advanced fuzzy and neural control*, Oulu, 16–17 Sept 2004, pp 67–72
- Angelov P, Xydeas C, Filev D (2004b) On-line identification of MIMO evolving Takagi–Sugeno fuzzy models. In: *Proceedings of the international joint conference on neural networks and international conference on fuzzy systems, IJCNN-FUZZ-IEEE*, Budapest, 25–29 July 2004, pp. 55–60
- Angelov P, Lughofer E, Klement PE (2005) Two approaches for data – driven design of evolving fuzzy systems: eTS and FLEXFIS. In: *Proceedings of the 2005 north American fuzzy information processing society, NAFIPS Annual Conference*, Ann Arbor, June 2005, pp 31–35
- Angelov P, Filev D, Kasabov N, Cordon O (eds) (2006) *Evolving fuzzy systems*. In: *Proceedings of the 2nd international symposium on evolving fuzzy systems, Ambleside*, 7–9 Sept 2006. pp 1–350
- Angelov P, Zhou X, Klawonn F (2007a) Evolving fuzzy rule-based classifiers. In: *Proceedings of the first 2007 IEEE international conference on computational intelligence applications for signal and image processing – a part of the IEEE symposium series on computational intelligence, SSCI-2007*, Honolulu, 1–5 April 2007, pp 220–225
- Angelov P, Zhou X, Lughofer E, Filev D (2007b) Architectures of evolving fuzzy rule-based classifiers. In: *Proceedings of the 2007 IEEE international conference on systems, man, and cybernetics*, Montreal, 7–10 Oct 2007, pp 2050–2055
- Angelov P, Kordon A, Zhou X (2008) Adaptive inferential sensors based on evolving fuzzy models: an industrial case study. *IEEE Trans Fuzzy Syst* (under review)
- Arulampalam MS, Maskell S, Gordon N (2002) A tutorial on particle filters for on-line nonlinear non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50(2): 174–188
- Astroem KJ, Wittenmark B (1994) *Adaptive control*. Prentice Hall, Upper Saddle River
- Azimi-Sadjadi MR, Yao D, Jamshidi AA, Dobeck GJ (2002) Underwater target classification in changing environments using an adaptive feature Mapping. *IEEE Trans Neural Netw* 13(5):1099–1111

- Badami VV, Chbat NW (1998) Home appliances get smart. *IEEE Spectr* 35(8):36–43
- Bezdek J (1974) Cluster validity with fuzzy sets. *J Cybern* 3(3):58–71
- Bonarini A, Lazaric A, Restelli M, Vitali P (2006) Self-development framework for reinforcement learning agents. In: Proceedings of the 5th international conference on development and learning, ICDL-06 New Delhi
- Carline D, Angelov PP, Clifford R (2005) Agile collaborative autonomous agents for robust underwater classification scenarios. In: The proceedings of the underwater defense technology conference, Amsterdam, June 2005
- Carpenter GA, Grossberg S (2003) Adaptive resonance theory. In: Arbib MA (ed) *The handbook of brain theory and neural networks*, 2nd edn. MIT Press, Cambridge, pp 87–90
- Chiu SL (1994) Fuzzy model identification based on cluster estimation. *J Intell Fuzzy Syst* 2:267–278
- Cordon O, Gomide F, Herrera F, Hoffmann F, Magdalena L (2004) Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets Syst* 141(1):5–31
- Domingos P, Hulten G (2001) Catching up with the data: research issues in mining data streams. In: Proceedings of the workshop on research issues in data mining and knowledge discovery, Santa Barbara
- Filev D, Tseng F (2006) Novelty detection-based machine health prognostics. In: Proceedings of the 2006 international symposium on evolving fuzzy systems. IEEE Press, USA, pp 193–199
- Filev D, Larson T, Ma L (2000) Intelligent control for automotive manufacturing – rule-based guided adaptation. In: Proceedings of the IEEE conference on industrial electronics, IECON-2000, Nagoya Oct 2000, pp 283–288
- Fortuna L, Graziani S, Rizzo A, Xibilia MG (2007) *Soft sensors for monitoring and control of industrial processes*. Springer, London
- Goldberg DE (1989) *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Reading
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning: data mining, inference and prediction*. Springer, Heidelberg
- Hornby AS (1974) *Oxford advance Learner's dictionary*. Oxford University Press, Oxford
- Huang G-B, Saratchandran P, Sundarajan N (2005) A generalized growing and pruning RBF (GGAP – RBF) neural network for function approximation. *IEEE Trans Neural Netw* 16(1):57–67
- Ishibuchi H, Nakashima T, Nii M (2004) *Classification and modeling with linguistic granules: advanced information processing*. Springer, Berlin
- Jones E, Angelov P, Xydeas C (2006) Recovery of LSP coefficients in VoIP systems using evolving Takagi–Sugeno fuzzy MIMO models. In: Proceedings of the 2006 international symposium on evolving fuzzy systems, Ambleside, 7–9 Sept 2006, pp 208–214
- Kailath T, Sayed AH, Hassibi B (2000) *Linear estimation*. Prentice Hall, Upper Saddle River
- Kalman RE (1960) A new approach to linear filtering and prediction problem. *Transactions of the American Society of Mechanical Engineering, ASME. Ser D J Basic Eng* 8:34–45
- Kanakakis V, Valavanis KP, Tsourveloudis NC (2004) Fuzzy-logic based navigation of underwater vehicles. *J Intell Robot Syst* 40:45–88
- Kasabov N (2001) Evolving fuzzy neural networks for on-line supervised/unsupervised, knowledge-based learning. *IEEE Trans Syst Man Cybern Syst* 31:902–918
- Kasabov N, Song Q (2002) DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans Fuzzy Syst* 10(2):144–154
- Klir G, Folger T (1988) *Fuzzy sets, uncertainty and information*. Prentice Hall, Englewood Cliffs
- Kohonen T (1995) *Self-organizing maps*, Series in Inf Sci, vol 30. Springer, Heidelberg
- Kordon A (2006) Inferential sensors as potential application area of intelligent evolving systems. 2006 international symposium on evolving fuzzy systems, Ambleside, 7–9 September 2006, key note presentation
- Kuncheva L (2000) *Fuzzy Classifiers*. Physica, Heidelberg
- Leng G, McGuinty TM, Prasad G (2005) An approach for on-line extraction of fuzzy rules using a self-organizing fuzzy neural network. *Fuzzy Sets Syst* 150(2):211–243
- Lin F-J, Lin C-H, Shen P-H (2001) Self-constructing fuzzy neural network speed controller for permanent-magnet synchronous motor drives. *IEEE Trans Fuzzy Syst* 9(5):751–759
- Liu PX, Meng MQ-X (2004) On-line data-driven fuzzy clustering with applications to real-time robotic tracking. *IEEE Trans Fuzzy Syst* 12(4):516–523
- Ljung L (1987) *System identification: theory for the user*. Prentice-Hall, New Jersey
- Lughofer E, Angelov P, Zhou X (2007) Evolving single- and multi-model fuzzy classifiers with FLEXFIS-class. In: Proceedings of the 2007 IEEE international conference on fuzzy systems, London, 23–26 July 2007, pp 363–368
- Macias J, Angelov P, Zhou X (2006) Predicting quality of the crude oil distillation using evolving Takagi–Sugeno fuzzymodels. In: Proceedings of the 2006 international symposium on evolving fuzzy systems, Ambleside, 7–9 Sept 2006, pp 201–207
- Macias-Hernandez JJ, Angelov P, Zhou X (2007) Soft sensor for predicting crude oil distillation side streams using Takagi Sugeno evolving fuzzy models. In: Proceedings of the 2007 IEEE international conference on systems, man, and cybernetics, Montreal, 7–10 Oct. 2007, pp 3305–3310
- Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man-Mach Stud* 7:1–13
- Massaro DW (1991) Integration versus interactive activation: the joint influence of stimulus and context in perception. *Cogn Psychol* 23:558–614

- Memon MA, Angelov P, Ahmed H (2006) An approach to real-time color-based object tracking. In: Proceedings 2006 international symposium on evolving fuzzy systems, Ambleside, 7–9 Sept 2006, pp 81–87
- Nauck D, Kruse R (1997) A neuro-fuzzy method to learn fuzzy classification rules from data. *Fuzzy Sets Syst* 89: 277–288
- Pioneer-3DX (2004) User Guide. ActiveMedia Robotics, Amherst
- Procyk TJ, Mamdani EH (1979) A linguistic self-organizing process controller. *Automatica* 15:15–30
- Psaltis D, Sideris A, Yamamura AA (1988) A multilayered neural network controller. *IEEE Trans Control Syst Manag* 8:17–21
- Qin SJ, Yue H, Dunia R (1997) Self-validating inferential sensors with application to air emission monitoring. *Ind Eng Chem Res* 36:1675–1685
- Setnes M, Roubos H (2000) Ga-fuzzy modeling and classification: complexity and performance. *IEEE Trans Fuzzy Syst* 8(5):509–522
- Shimajima K, Fukuda T, Hashegawa Y (1995) Self-tuning modeling with adaptive membership function, rules, and hierarchical structure based on genetic algorithm. *Fuzzy Sets Syst* 71:295–309
- Sifalakis M, Hutchison D (2004) From active networks to cognitive networks. In: Proceedings of the ICRC Dagstuhl seminar 04411 on service management and self-organization in IP-based networks, Waden. October 2004
- Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. *IEEE Trans Syst Man Cybern B – Cybern* 15:116–132
- Valavanis K (2006) Unmanned vehicle navigation and control: a fuzzy logic perspective. In: Proceedings of the 2006 international symposium on evolving fuzzy systems. Ambleside, 7–9 Sept. 2006, pp 200–207
- Vapnik VN (1998) The statistical learning theory. Springer, Berlin
- Wang L-X (1992) Fuzzy systems are universal approximators. In: Proceedings of the first IEEE international conference on fuzzy systems, FUZZ-IEEE – 1992, San Diego, pp 1163–1170
- Widrow B, Stearns S (1985) Adaptive signal processing. Prentice Hall, Englewood Cliffs
- Xydeas C, Angelov P, Chiao S, Reoullas M (2006) Advances in EEG signals classification via Dependant HMM models and evolving fuzzy classifiers. *Int J Comput Biol Medicine, special issue on Intell Technol Bio-Inform Medicine* 36(10):1064–1083
- Yager R (2006) Learning methods for intelligent evolving systems. In: Proceedings 2006 international symposium on evolving fuzzy systems. Ambleside, 7–9 Sept. 2006, pp 3–7
- Yager RR, Filev DP (1993) Learning of fuzzy rules by mountain clustering. In: Proceedings of the SPIE Conference on application of fuzzy logic technology, Boston, pp 246–254
- Yager RR, Filev DP (1994) Essentials of fuzzy modeling and control. Wiley, New York
- Zadeh LA (1993) Soft computing. Introductory lecture for the 1st European congress on fuzzy and intelligent technologies EUFIT'93, Aachen, pp vi–vii
- Zhou X-W, Angelov P (2006) Real-time joint landmark recognition and classifier generation by an evolving fuzzy system. In: Proceedings of the 2006 IEEE world congress on computational intelligence, WCCI-2006, Vancouver. 16–21 July 2006, pp 6314–6321
- Zhou X, Angelov P (2007) An approach to autonomous self-localization of a mobile robot in completely unknown environment using evolving fuzzy rule-based classifier. In: Proceedings of the First 2007 IEEE Int Symposium on Computational Intelligence Applications for Defense and Security – a part of the IEEE Symposium Series on Computational Intelligence, SSCI-2007, Honolulu, 1–5 April 2007, pp 131–138

Books and Reviews

- Angelov P, Xydeas C (2006) Fuzzy systems design: direct and indirect approaches. *Int J Soft Comput, special issue on New Trends in Fuzzy Modeling part I: Novel Approaches* 10(9):836–849
- Angelov P, Zhou X (2006) Evolving fuzzy systems from data streams in real-time. In: Proceedings 2006 international symposium on evolving fuzzy systems, Ambleside, 7–9 Sept. 2006, pp 29–35
- Bentley PJ (2000) Evolving fuzzy detectives: an investigation into the evolution of fuzzy rules. In: Suzuki, Roy, Ovask, Furuhashi, Dote (eds) *Soft computing in industrial applications*. Springer, London
- Chan Z, Kasabov N (2004) Evolutionary computation for on-line and off-line parameter tuning of evolving fuzzy neural networks. *Int J Comput Intell Appl* 4(3): 309–319
- Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery: an overview, advances in knowledge discovery and data mining. MIT Press, Boston
- Fritzke B (1994) Growing cell structures – a self-organizing network for unsupervised and supervised learning. *Neural Netw* 7(9):1441–1460
- Futschik M, Reeve A, Kasabov N (2003) Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue. *Artif Intell Med* 28: 165–189
- Hopner F, Klawonn F (2000) Obtaining interpretable fuzzy models from fuzzy clustering and fuzzy regression. In: Proceedings of the 4th intern Conf on knowledge-based intelligent engineering systems (KES). Brighton, pp 162–165
- Huang G-B, Saratchandran P, Sundarajan N (2005a) A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. *IEEE Trans Neural Netw* 16(1):57–67
- Huang L, Song Q, Kasabov N (2005b) Evolving connectionist systems based role allocation of robots for soccer playing. In: Proceedings of the joint 2005 international symposium on intelligent control

- and 13th Mediterranean conference on control and automation, ISIC – MED – 2005, Limassol. 27–29 June 2005
- Jang JSR (1993) ANFIS: adaptive network-based fuzzy inference systems. *IEEE Trans on Syst Man Cybernt B – Cybern* 23(3):665–685
- Juang C-F, Lin X-T (1999) A recurrent self-organizing neural fuzzy inference network. *IEEE Trans Neural Netw* 10:828–845
- Kasabov N (2006a) Adaptation and interaction in dynamical systems: modelling and rule discovery through evolving connectionist systems. *Appl Soft Comput* 6(3):307–322
- Kasabov N (2006b) Evolving connectionist systems: brain-, gene-, and, quantum inspired computational intelligence. Springer, London
- Kasabov N, Chan Z, Song Q, Greer D (2005) Evolving neuro-fuzzy systems with evolutionary parameter self-optimisation. In: Do adaptive smart systems exist? Series study in fuzziness, vol 173. Physica, Heidelberg
- Kim K, Baek J, Kim E, Park M (2005) TSK fuzzy model based on-line identification. In: Proceedings of the 11th international fuzzy systems association, IFSA world congress, Beijing, pp 1435–1439
- Klinkenberg R, Joachims T (2000) Detection concept drift with support vector machines. Proceedings of the 7th international conference on machine learning (ICML). Morgan Kaufman, Stanford University, pp. 487–494
- Marin-Blazquez JG, Shen Q (2002) From approximative to descriptive fuzzy classifiers. *IEEE Trans Fuzzy Syst* 10(4):484–497
- Marshall MR, Song Q, Ma TM, MacDonell S, Kasabov N (2005) Evolving connectionist system versus algebraic formulae for prediction of renal function from serum creatinine. *Kidney Int* 6:1944–1954
- Ozawa S, Pang S, Kasabov N (2004) A modified incremental principal component analysis for on-line learning of feature space and classifier. In: Lecture notes in artificial intelligence LNAI, vol 3157. Springer, Berlin, pp 231–240
- Pang S, Ozawa S, Kasabov N (2005) Incremental linear discriminant analysis for classification of data streams. *IEEE Trans Syst Man Cybern B – Cybern* 35(5): 905–914
- Plat J (1991) A resource allocation network for function interpolation. *Neural Comput* 3(2):213–225
- Widmer G, Kubat M (1996) Learning in the presence of concept drift and hidden contexts. *Mach Learn* 23(1): 69–101

Fuzzy Logic, Type-2 and Uncertainty

Robert I. John¹ and Jerry M. Mendel²

¹Centre for Computational Intelligence, School of Computing, De Montfort University, Leicester, UK

²Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Type-2 Fuzzy Systems

Generalized Type-2 Fuzzy Systems

Interval Type-2 Fuzzy Sets and Systems

Future Directions

Bibliography

Glossary

Type-1 fuzzy sets Are the underlying component in fuzzy logic where uncertainty is represented by a number between one and zero.

Type-2 fuzzy sets Are where the uncertainty is represented by a type-1 fuzzy set.

Interval type-2 fuzzy sets Are where the uncertainty is represented by a type-1 fuzzy set where the membership grades are unity.

Definition of the Subject

Type-2 fuzzy logic was first defined in 1975 by Zadeh and is an increasingly popular area for research and applications. The reason for this is because it appears to tackle the fundamental problem with type-1 fuzzy logic in that it is unable to handle the many uncertainties in real systems. Type-2 fuzzy systems are conceptually and

mathematically more difficult to understand and implement but the proven applications show that the effort is worth it and type-2 fuzzy systems are at the forefront of fuzzy logic research and applications. These systems rely on the notion of a type-2 fuzzy set where the membership grades are type-1 fuzzy sets.

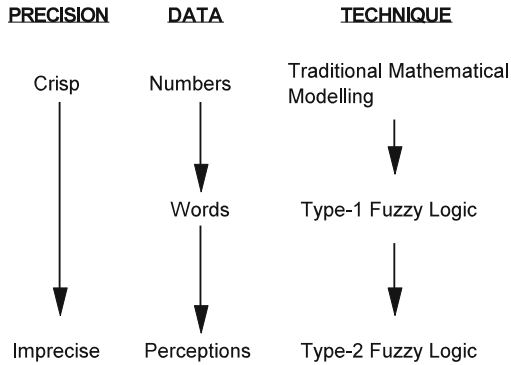
Introduction

Fuzzy sets (Bustince 2000) have, over the past forty years, laid the basis for a successful method of modeling uncertainty, vagueness and imprecision in a way that no other technique has been able. The use of fuzzy sets in real computer systems is extensive, particularly in consumer products and control applications.

Fuzzy logic (a logic based on fuzzy sets) is now a mainstream technique in everyday use across the world. The number of applications is many, and growing, in a variety of areas, for example, heat exchange, warm water pressure, aircraft flight control, robot control, car speed control, power systems, nuclear reactor control, fuzzy memory devices and the fuzzy computer, control of a cement kiln, focusing of a camcorder, climate control for buildings, shower control and mobile robots. The use of fuzzy logic is not limited to control. Successful applications, for example, have been reported in train scheduling, system modeling, computing, stock tracking on the Nikkei stock exchange and information retrieval.

Type-1 fuzzy sets represent uncertainty using a number in $[0, 1]$ whereas type-2 fuzzy sets represent uncertainty by a function. This is discussed in more detail later in the article. Essentially, the more imprecise or vague the data is, then type-2 fuzzy sets offer a significant improvement on type-1 fuzzy sets. Figure 1 shows the view taken here of the relationships between levels of imprecision, data and technique.

As the level of imprecision increases then type-2 fuzzy logic provides a powerful paradigm



Fuzzy Logic, Type-2 and Uncertainty,
Fig. 1 Relationships between imprecision, data and fuzzy technique

for potentially tackling the problem. Problems that contain crisp, precise data do not, in reality, exist. However some problems can be tackled effectively using mathematical techniques where the assumption is that the data is precise. Other problems (for example, in control) use imprecise terminology that can often be effectively modeled using type-1 fuzzy sets. Perceptions, it is argued here, are at a higher level of imprecision and type-2 fuzzy sets can effectively model this imprecision.

The reason for this lies in some of the problems associated with type-1 fuzzy logic systems. Although successful in the control domain they have not delivered as well in systems that attempt to replicate human decision making. It is our view that this is because a type-1 fuzzy logic system (FLS) has some uncertainties which cannot be modeled properly by type-1 fuzzy logic. The sources of the uncertainties in type-1 FLSs are:

- The meanings of the words that are used in the antecedents and consequents of rules can be uncertain (words mean different things to different people).
- Consequents may have a histogram of values associated with them, especially when knowledge is extracted from a group of experts who do not all agree.
- Measurements that activate a type-1 FLS may be noisy and therefore uncertain.

- The data that are used to tune the parameters of a type-1 FLS may also be noisy.

The uncertainties described all essentially have to do with the uncertainty contained in a type-1 fuzzy set. A type-1 fuzzy set can be defined in the following way:

Let X be a universal set defined in a specific problem, with a generic element denoted by x . A fuzzy set A in X is a set of ordered pairs:

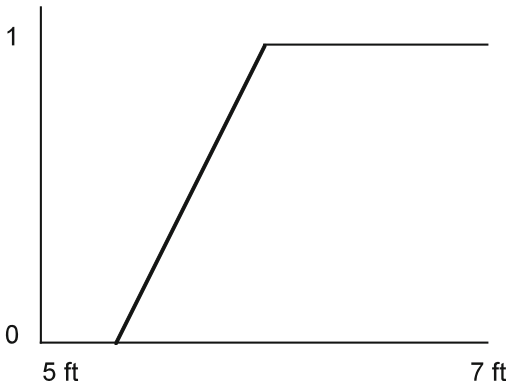
$$A = \{(x, \mu_A(x) \mid x \in X)\},$$

where $\mu_A : X \rightarrow [0, 1]$ is called the membership function A of and $\mu_A(x)$ represents the degree of membership of the element x in A .

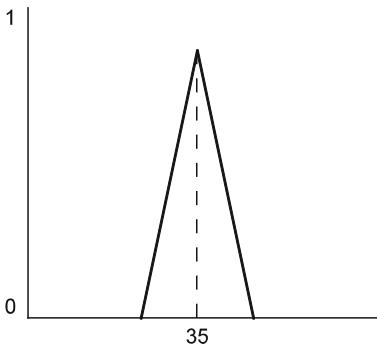
The key points to draw from this definition of a fuzzy set are:

- The members of a fuzzy set are members to some degree, known as a *membership grade* or degree of membership.
- A fuzzy set is fully determined by the membership function.
- The membership grade is the degree of belonging to the fuzzy set. The larger the number (in $[0, 1]$) the more the degree of belonging.
- The translation from x to $\mu_A(x)$ is known as *fuzzification*.
- A fuzzy set is either continuous or discrete.
- Graphical representation of membership functions is very useful. For example, the fuzzy set 'Tall' might be represented as shown in Fig. 2 where someone who is of height five feet has a membership grade of zero while someone who is of height seven feet is tall to degree one, with heights in between having membership grade between one and zero. The example shown is linear but, of course, it could be any function.

Fuzzy sets offer a practical way of modeling what one might refer to as 'fuzziness'. The real world can be characterized by the fact that much of it is imprecise in one form or other. For a clear exposition (important to the notion of, and argument for, type-2 sets) two ideas of 'fuzziness' can be considered important – imprecision and vagueness (linguistic uncertainty).



Fuzzy Logic, Type-2 and Uncertainty, Fig. 2 The fuzzy set 'Tall'



Fuzzy Logic, Type-2 and Uncertainty, Fig. 3 The fuzzy number 'About 35'

Imprecision

As has already been discussed, in many physical systems measurements are never precise (a physical property can always be measured more accurately). There is imprecision inherent in measurement. Fuzzy numbers are one way of capturing this imprecision by having a fuzzy set represent a real number where the numbers in an interval near to the number are in the fuzzy set to some degree. So, for example, the fuzzy number 'About 35' might look like the fuzzy set in Fig. 3 where the numbers closer to 35 have membership nearer unity than those that are further away from 35.

Vagueness or Linguistic Uncertainty

Another use of fuzzy sets is where words have been used to capture imprecise notions, loose concepts or perceptions. We use words in our everyday language that we, and the intended

audience, know what we want to convey but the words cannot be precisely defined. For example, when a bank is considering a loan application somebody may be assessed as a good risk in terms of being able to repay the loan. Within the particular bank this notion of a good risk is well understood. It is not a black and white decision as to whether someone is a good risk or not – they are a good risk *to some degree*.

Type-2 Fuzzy Systems

A type-1 fuzzy system uses type-1 fuzzy sets in either the antecedent and/or the consequent of type-1 fuzzy if-then rules and a type-2 fuzzy system deploys type-2 fuzzy sets in either the antecedent and/or the consequent of type-2 fuzzy rules.

Fuzzy systems usually have the following features:

- The *fuzzy sets* as defined by their membership functions. These fuzzy sets are the basis of a fuzzy system. They capture the underlying properties or knowledge in the system.
- The *if-then rules* that combine the fuzzy sets – in a rule set or knowledge base.
- The fuzzy *composition* of the rules. Any fuzzy system that has a set of if-then rules has to combine the rules.
- Optionally, the defuzzification of the solution fuzzy set. In many (most) fuzzy systems there is a requirement that the final output be a 'crisp' number. However, for certain fuzzy paradigms the output of the system is a fuzzy set, or its associated word. This solution set is 'defuzzified' to arrive at a number.

Type-1 fuzzy sets are, in fact, crisp and not at all fuzzy, and are two dimensional. A domain value x is simply represented by a number in $[0, 1]$ – the membership grade. The methods for combining type-1 fuzzy sets in rules are also precise in nature.

Type-2 fuzzy sets, in contrast, are three dimensional. The membership grade for a given value in a type-2 fuzzy set is a type-1 fuzzy set. A formal

definition of a type-2 fuzzy set is given in the following:

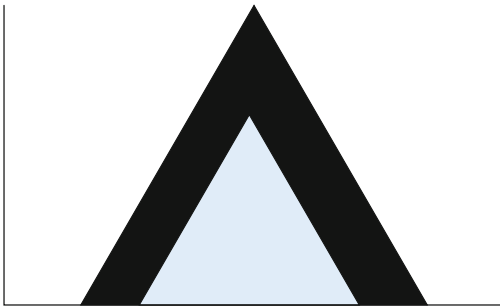
A type-2 fuzzy set, $\tilde{A}$, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x, u)$, where $x \in X$ and $u \in J_x$ subset $JM [0, 1]$, and J_x is called the primary membership, i. e.

$$\tilde{A} = \left\{ \left((x, u), \mu_{\tilde{A}}(x, u) \right) \mid \forall x \in X, \right. \\ \left. \forall J_x \text{ subset } JM [0, 1] \right\}. \quad (1)$$

A useful way of looking at a type-2 fuzzy set is by considering its Footprint of Uncertainty (FOU). This is a two dimensional view of a type-2 fuzzy set. See Fig. 4 for a simple example. The shaded area represents the Union of all the J_x .

An effective way to compare type-1 fuzzy sets and type-2 fuzzy sets is by use of a simple example. Suppose, for a particular application, we wish to describe the imprecise concept of ‘tallness’. One approach would be to use a type-1 fuzzy set $tall_1$. Now suppose we are only considering three members of this set – Michael Jordan, Danny Devito and Robert John. For the type-1 fuzzy approach one might say that Michael Jordan is $tall_1$ to degree 0.95, Danny Devito to degree 0.4 and Robert John to degree 0.6. This can be written as

$$tall_1 = 0.95/Michael\ Jordan \\ + 0.4/Danny\ Devito \\ + 0.6/Robert\ John.$$



Fuzzy Logic, Type-2 and Uncertainty, Fig. 4 A typical FOU of a type-2 set

A type-2 fuzzy set ($tall_2$) that models the concept of ‘tallness’ could be

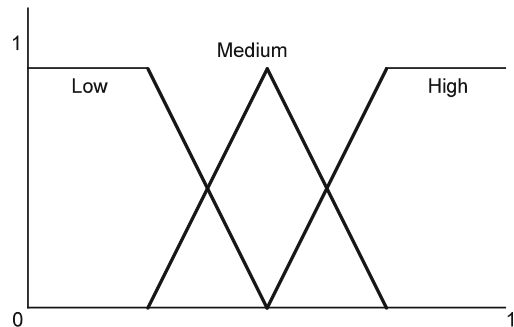
$$tall_2 = High_1/Michael\ Jordan \\ + Low_1/Danny\ Devito \\ + Medium_1/Robert\ John,$$

where $High_1$, Low_1 and $Medium_1$ are type-1 fuzzy sets.

Figure 5 shows what the sets $High_1$, Low_1 and $Medium_1$ might look like if represented graphically. As can be seen the x axis takes values between 0 and 1, as does the y axis (μ).

Type-1 sets have an x axis representing the domain –in this case the height of an individual. Type-2 sets employ type-1 sets as the membership grades. Therefore, these fuzzy sets of type-2 allow for the idea that a fuzzy approach does not necessarily have membership grades $[0, 1]$ in but the degree of membership for the member is itself a type-1 fuzzy set. As can be seen, by the simple example, there is an inherent extra fuzziness offered by type-2 fuzzy sets over and above a type-1 approach. So, a type-2 fuzzy set could be called a fuzzy-fuzzy set.

Real situations do not allow for precise numbers in $[0, 1]$. In a control application, for instance, can we say that a particular temperature, t , belongs to the type-1 fuzzy set hot_1 with a membership grade x precisely? No. Firstly it is highly likely that the membership could just as well be $x - 0.1$ for example. Different experts would attach different membership grades and, indeed, the same expert might well give different values on different days! On top of this



Fuzzy Logic, Type-2 and Uncertainty, Fig. 5 The Fuzzy Sets $High_1$, Low_1 and $Medium_1$

uncertainty there is always some uncertainty in the measurement of t . So, we have a situation where an uncertain measurement is matched *precisely* to another uncertain value!! Type-2 fuzzy sets on the other hand, for certain appropriate applications, allow for this uncertainty to be modeled by not using precise membership grades but imprecise type-1 fuzzy sets.

So that type-2 sets can be used in a fuzzy system (in a similar manner to type-1 fuzzy sets) a method is required for computing the intersection (AND) and union (OR) of two type-2 sets. Suppose we have two type-2 fuzzy sets, $\tilde{A}$ and $\tilde{B}$ in X and $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$ are two secondary membership functions of $\tilde{A}$ and $\tilde{B}$ respectively, represented as:

$$\begin{aligned}\mu_{\tilde{A}}(x) &= f(u_1)/u_1 + f(u_2)/u_2 + \dots + f(u_n)/u_n \\ &= \sum_i f(u_i)/u_i, \\ \mu_{\tilde{B}}(x) &= g(w_1)/w_1 + g(w_2)/w_2 + \dots + g(w_m)/w_m \\ &= \sum_j g(w_j)/w_j,\end{aligned}$$

where the functions f and g are membership functions of fuzzy grades and $\{u_i, i = 1, 2, \dots, n\}$, $\{w_j, j = 1, 2, \dots, m\}$, are the members of the fuzzy grades.

Union of Type-2 Fuzzy Sets

The union ($\cup$) of two type-2 fuzzy sets ($\tilde{A}, \tilde{B}$) corresponding to $\tilde{A} \text{ OR } \tilde{B}$ is given by:

$$\begin{aligned}\tilde{A} \cup \tilde{B} &\Leftrightarrow \mu_{\tilde{A} \cup \tilde{B}}(x) = \tilde{A} \text{ join } \tilde{B} \\ &= \sum_{ij} \frac{(f(u_i)JMg(w_j))}{(u_iJMw_j)}.\end{aligned}$$

Intersection of Type-2 Fuzzy Sets

The intersection ($\cap$) of two type-2 fuzzy sets ($\tilde{A}, \tilde{B}$) corresponding to $\tilde{A} \text{ AND } \tilde{B}$ is given by:

$$\begin{aligned}\tilde{A} \cap \tilde{B} &\Leftrightarrow \mu_{\tilde{A} \cap \tilde{B}}(x) = \tilde{A} \text{ meet } \tilde{B} \\ &= \sum_{ij} \frac{(f(u_i)JMg(w_j))}{(u_iJMw_j)},\end{aligned}$$

where $\text{join } JM$ denotes *join* and $\text{meet } JM$ denotes *meet*.

Thus the join and meet allow for us to combine type-2 fuzzy sets for the situation where we wish to 'AND' or 'OR' two type-2 fuzzy sets. Join and meet are the building blocks for type-2 fuzzy relations and type-2 fuzzy inferencing with type-2 if-then rules.

Type-2 fuzzy if-then rules (type-2 rules) are similar to type-1 fuzzy if-then rules. An example type-2 if-then rule is given by

$$\text{IF } x \text{ is } \tilde{A} \text{ and } y \text{ is } \tilde{B} \text{ then } z \text{ is } \tilde{C}. \quad (2)$$

Obviously the rule could have a more complex antecedent connected by AND. Also the consequent of the rule could be type-1 or, indeed, crisp.

Type-2 output processing can be done in a number of ways through type-reduction (e. g. Centroid, Centre of Sums, Height, Modified Height and Center-of-Sets) that produces a type-1 fuzzy set, followed by defuzzification of that set.

Generalized Type-2 Fuzzy Systems

So far we have been discussing type-2 fuzzy systems where the secondary membership function can take any form – these are known as *generalized* type-2 fuzzy sets. Historically these have been difficult to work with because the complexity of the calculations is too high for real applications. Recent developments (Coupland 2007; Coupland and John 2005, 2007) mean that it is now possible to develop type-2 fuzzy systems where, for example, the secondary membership functions are triangular in shape. This is relatively new but offers an exciting opportunity to capture the uncertainty in real applications. We expect the interest in this area to grow considerably but for the purposes of this article we will concentrate on the detail of interval type-2 fuzzy systems where the secondary membership function is always unity.

Interval Type-2 Fuzzy Sets and Systems

As of this date, interval type-2 fuzzy sets (IT2 FSs) and interval type-2 fuzzy logic systems

(IT2 FLSs) are most widely used because it is easy to compute using them. IT2 FLSs are also known as *interval-valued FLSs* for which there is a very extensive literature (e.g., (Bustince 2000)), see the many references in this article, (Gorzalczany 1987, 1988; Mendel 2007a; Zadeh 1975)). This section focuses on IT2 FLSs and IT2 FLSs.

Interval Type-2 Fuzzy Sets

An IT2 FS $\tilde{A}$ is characterized as (much of the background material in this sub-section is taken from (Mendel et al. 2006b); see also (Mendel 2007b)):

$$\begin{aligned}\tilde{A} &= \int_{x \in X} \int_{u \in J_x \subseteq [0,1]} 1/(x, u) \\ &= \int_{x \in X} \left[\int_{u \in J_x \subseteq [0,1]} 1/u \right] /x,\end{aligned}\quad (3)$$

where x , the *primary variable*, has domain X ; $u \in U$, the *secondary variable*, has domain J_x at each $x \in X$; J_x is called the *primary membership* of x and is defined below in (9); and, the *secondary grades* of $\tilde{A}$ all equal 1. Note that for continuous X and U , (3) means: $\tilde{A} : X \rightarrow \{[a, b] : 0 \leq a \leq b \leq 1\}$.

The bracketed term in (3) is called the *secondary MF*, or *vertical slice*, of $\tilde{A}$, and is denoted $\mu_{\tilde{A}}^{\sim}(x)$, i. e.

$$\mu_{\tilde{A}}^{\sim}(x) = \int_{u \in J_x \subseteq [0,1]} 1/u, \quad (4)$$

so that $\tilde{A}$ can be expressed in terms of its vertical slices as:

$$\tilde{A} = \int_{x \in X} \mu_{\tilde{A}}^{\sim}(x) /x. \quad (5)$$

Uncertainty about $\tilde{A}$ is conveyed by the union of all the primary memberships, which is called the *footprint of uncertainty* (FOU) of $\tilde{A}$ (see Fig. 6), i. e.

$$\begin{aligned}\text{FOU}(\tilde{A}) &= \bigcup_{x \in X} J_x \\ &= \{(x, u) : u \in J_x \subseteq [0, 1]\}.\end{aligned}\quad (6)$$

The *upper membership function* (UMF) and *lower membership function* (LMF) of $\tilde{A}$ are two type-1 MFs that bound the FOU (Fig. 6). $\text{UMF}(\tilde{A})$ is associated with the upper bound of $\text{FOU}(\tilde{A})$ and is denoted $\bar{\mu}_{\tilde{A}}^{\sim}(x), \forall x \in X$, and $\text{LMF}(\tilde{A})$ is associated with the lower bound of $\text{FOU}(\tilde{A})$ and is denoted $\underline{\mu}_{\tilde{A}}^{\sim}(x), \forall x \in X$, i. e.

$$\text{UMF}(\tilde{A}) \equiv \bar{\mu}_{\tilde{A}}^{\sim}(x) = \overline{\text{FOU}(\tilde{A})} \quad \forall x \in X, \quad (7)$$

$$\text{LMF}(\tilde{A}) \equiv \underline{\mu}_{\tilde{A}}^{\sim}(x) = \underline{\text{FOU}(\tilde{A})} \quad \forall x \in X. \quad (8)$$

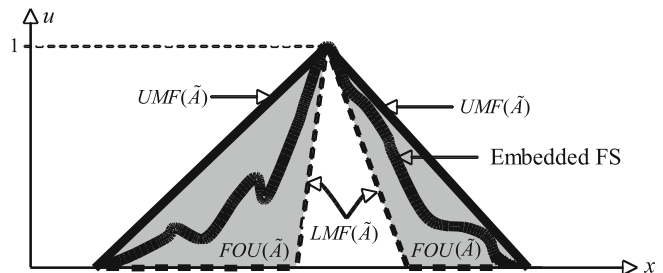
Note that J_x is an *interval set*, i. e.

$$J_x = [\underline{\mu}_{\tilde{A}}^{\sim}(x), \bar{\mu}_{\tilde{A}}^{\sim}(x)]. \quad (9)$$

This set is discrete when U is discrete and is continuous when U is continuous. Using (9), the $\text{FOU}(\tilde{A})$ in (6) can also be expressed as

$$\text{FOU}(\tilde{A}) = \bigcup_{x \in X} [\underline{\mu}_{\tilde{A}}^{\sim}(x), \bar{\mu}_{\tilde{A}}^{\sim}(x)]. \quad (10)$$

Fuzzy Logic, Type-2 and Uncertainty, Fig. 6 FOU (shaded), LMF (dashed), UMF (solid) and an embedded FS (wavy line) for IT2 FS $\tilde{A}$



A very compact way to describe an IT2 FS is (Mendel et al. 2006a):

$$\tilde{A} = 1/\text{FOU}(\tilde{A}), \quad (11)$$

where this notation means that the secondary grade equals 1 for all elements of $\text{FOU}(\tilde{A})$.

For continuous universes of discourse X and U , an *embedded* IT2 FS $\tilde{A}_e$ is

$$\tilde{A}_e = \int_{x \in X} [1/u]/x \quad u \in J_x. \quad (12)$$

Note that (12) means: $\tilde{A}_e : X \rightarrow \{u : 0 \leq u \leq 1\}$. The set $\tilde{A}_e$ is embedded in $\tilde{A}$ such that at each x it only has one secondary variable (i.e., one primary membership whose secondary grade equals 1). Examples of $\tilde{A}_e$ are $1/\underline{\mu}_A(x)$ and $1/\bar{\mu}_A(x)$, $\forall x \in X$. In this notation it is understood that the secondary grade equals 1 at all elements in $\underline{\mu}_A(x)$ or $\bar{\mu}_A(x)$.

For discrete universes of discourse X and U , in which x has been discretized into N values and at each of these values u has been discretized into M_i values, an *embedded* IT2 FS $\tilde{A}_e$ has N elements, where $\tilde{A}_e$ contains exactly one element from $J_{x_1}, J_{x_2}, \dots, J_{x_N}$, namely $u_1, u_2, \dots, u_N$, each with a secondary grade equal to 1, i. e.,

$$\tilde{A}_e = \sum_{i=1}^N [1/u_i]/x_i, \quad (13)$$

where $u_i \in J_{x_i}$. Set $\tilde{A}_e$ is embedded in $\tilde{A}$, and, there are a total of $n_A = \prod_{i=1}^N M_i$ embedded T2 FSs.

Associated with each $\tilde{A}_e$ is an *embedded* T1 FS A_e , where

$$A_e = \int_{x \in X} u/x \quad u \in J_x. \quad (14)$$

The set A_e , which acts as the domain for $\tilde{A}_e$ (i. e., $\tilde{A}_e = 1/A_e$) is the union of all the primary memberships of the set $\tilde{A}_e$ in (12). Examples of A_e are $\underline{\mu}_A(x)$ and $\bar{\mu}_A(x)$, $\forall x \in X$.

When the universes of discourse X and U are continuous then there is an uncountable number of embedded IT2 and T1 FSs in $\tilde{A}$. Because such sets are only used for theoretical purposes and are

not used for computational purposes, this poses no problem.

For discrete universes of discourse X and U , an *embedded* T1 FS A_e has N elements, one each from $J_{x_1}, J_{x_2}, \dots, J_{x_N}$, namely $u_1, u_2, \dots, u_N$, i. e.,

$$A_e = \sum_{i=1}^N u_i/x_i. \quad (15)$$

Set A_e is the union of all the primary memberships of set A_e and, there are a total of $\prod_{i=1}^N M_i$ *embedded* T1 FSs.

Theorem 1 (Representation Theorem (RT) (Mendel and John 2002) Specialized to an IT2 FS (Mendel et al. 2006a)) For an IT2 FS, for which X and U are discrete, $\tilde{A}$ is the union of all of its embedded IT2 FSs. Equivalently, the domain of $\tilde{A}$ is equal to the union of all of its embedded T1 FSs, so that $\tilde{A}$ can be expressed as

$$\begin{aligned} \tilde{A} &= \sum_{j=1}^{n_A} \tilde{A}_e^j = 1/\text{FOU}(\tilde{A}) = 1/\sum_{j=1}^{n_A} A_e^j \\ &= 1/\sum_{i=1}^N u_i^j/x_i \\ &= 1/\bigcup_{\forall x \in X} \left\{ \underline{\mu}_A(x), \dots, \bar{\mu}_A(x) \right\} \end{aligned} \quad (16)$$

and if X is a continuous universe, then the infinite set $\left\{ \underline{\mu}_A(x), \dots, \bar{\mu}_A(x) \right\}$ is replaced by the interval set $\left[\underline{\mu}_A(x), \bar{\mu}_A(x) \right]$.

This RT is arguably the most important result in IT2 FS theory because it can be used as the starting point for solving all problems involving IT2 FSs. It expresses an IT2 FS in terms of T1 FSs, so that all results for problems that use IT2 FSs can be solved using T1 FS mathematics (Mendel et al. 2006a). Its use leads to the structure of the solution to a problem, after which efficient computational methods must be found to implement that structural solution. Table 1 summarizes set theoretic operations and uncertainty measures all of which were computed using the RT. Additional results for similarity measures are in (Wu and Mendel 2008).

Fuzzy Logic, Type-2 and Uncertainty, Table 1 Results for IT2 FSs

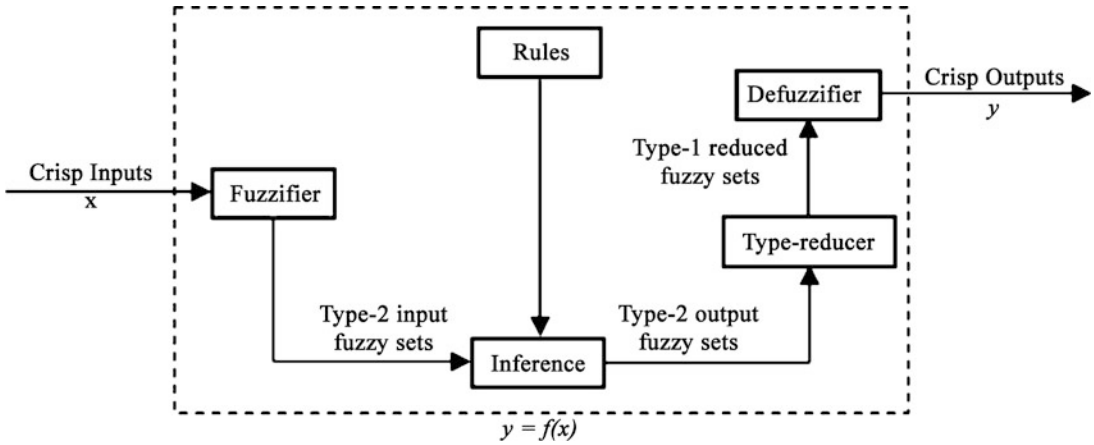
Set theoretic operations (Mendel et al. 2006a)	
Union	$\tilde{A} \cup \tilde{B} = 1 / \cup_{x \in X} \left[\underline{\mu}_{\tilde{A}}(x) \vee \underline{\mu}_{\tilde{B}}(x), \overline{\mu}_{\tilde{A}}(x) \vee \overline{\mu}_{\tilde{B}}(x) \right]$
Intersection	$\tilde{A} \cap \tilde{B} = 1 / \cup_{x \in X} \left[\underline{\mu}_{\tilde{A}}(x) \wedge \underline{\mu}_{\tilde{B}}(x), \overline{\mu}_{\tilde{A}}(x) \wedge \overline{\mu}_{\tilde{B}}(x) \right]$
Complement	$\tilde{\tilde{A}} = 1 / \cup_{x \in X} \left[1 - \underline{\mu}_{\tilde{A}}(x), 1 - \overline{\mu}_{\tilde{A}}(x) \right]$
Uncertainty measures (Karnik and Mendel 2001; Mendel and Wu 2007; Wu and Mendel 2007)	
Centroid	$C_{\tilde{A}} = \left[c_l(\tilde{A}), c_r(\tilde{A}) \right] = \left[\frac{\sum_{i=1}^L x_i \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^N x_i \underline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^L \underline{\mu}_{\tilde{A}}(x_i) + \sum_{i=L+1}^N \underline{\mu}_{\tilde{A}}(x_i)}, \frac{\sum_{i=1}^R x_i \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^N x_i \overline{\mu}_{\tilde{A}}(x_i)}{\sum_{i=1}^R \overline{\mu}_{\tilde{A}}(x_i) + \sum_{i=R+1}^N \overline{\mu}_{\tilde{A}}(x_i)} \right]$ L and R computed using the KM algorithms in Table 2.
Cardinality	$P_{\tilde{A}} = \left[p_l(\tilde{A}), p_r(\tilde{A}) \right] = \left[p(\underline{\mu}_{\tilde{A}}(x)), p(\overline{\mu}_{\tilde{A}}(x)) \right], p(B) = X \sum_{i=1}^N \mu_B(x_i) / N$
Fuzziness	$F_{\tilde{A}} = \left[f_1(\tilde{A}), f_2(\tilde{A}) \right] = [f_1(A_{e1}), f_2(A_{e2})], f(A) = h\left(\sum_{i=1}^N g(\mu_A(x_i))\right) A_{e1} : \mu_{A_{e1}}(x) =$ $\begin{cases} \overline{\mu}_{\tilde{A}}(x) & \overline{\mu}_{\tilde{A}}(x) \text{ is further away from 0.5 than } \underline{\mu}_{\tilde{A}}(x) \\ \underline{\mu}_{\tilde{A}}(x) & \text{otherwise} \end{cases} A_{e2} : \mu_{A_{e2}}(x) =$ $\begin{cases} \overline{\mu}_{\tilde{A}}(x) & \text{both } \overline{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are below 0.5} \\ \underline{\mu}_{\tilde{A}}(x) & \text{both } \overline{\mu}_{\tilde{A}}(x) \text{ and } \underline{\mu}_{\tilde{A}}(x) \text{ are above 0.5} \\ 0.5 & \text{otherwise} \end{cases}$
Variance	$V_{\tilde{A}} = \left[v_l(\tilde{A}), v_r(\tilde{A}) \right] = \left[\min_{\forall A_e} v_{\tilde{A}}(A_e), \max_{\forall A_e} v_{\tilde{A}}(A_e) \right] v_{\tilde{A}}(A_e) =$ $\sum_{i=1}^N \left[x_i - c(\tilde{A}) \right]^2 \mu_{A_e}(x_i) / \sum_{i=1}^N \mu_{A_e}(x_i), c(\tilde{A}) = \left[c_l(\tilde{A}) + c_r(\tilde{A}) \right] / 2$ KM algorithms are used to compute $v_l(\tilde{A})$ and $v_r(\tilde{A})$
Skew	$S_{\tilde{A}} = \left[s_l(\tilde{A}), s_r(\tilde{A}) \right] = \left[\min_{\forall A_e} s_{\tilde{A}}(A_e), \max_{\forall A_e} s_{\tilde{A}}(A_e) \right] s_{\tilde{A}}(A_e) =$ $\sum_{i=1}^N \left[x_i - c(\tilde{A}) \right]^3 \mu_{A_e}(x_i) / \sum_{i=1}^N \mu_{A_e}(x_i), c(\tilde{A}) = \left[c_l(\tilde{A}) + c_r(\tilde{A}) \right] / 2$ KM algorithms are used to compute $s_l(\tilde{A})$ and $s_r(\tilde{A})$

Interval Type-2 Fuzzy Logic Systems

An interval type-2 fuzzy logic system (IT2 FLS), which is a FLS that uses at least one IT2 FS, contains five components – fuzzifier, rules, inference engine, type-reducer and defuzzifier – that are inter-connected, as shown in Fig. 7 (the background material in this sub-section is taken from (Mendel et al. 2006b)). The IT2 FLS can be viewed as a mapping from inputs to outputs (the path in Fig. 7, from “Crisp Inputs” to “Crisp Outputs”), and this mapping can be expressed quantitatively as $y = f(x)$, and is also known as *interval type-2 fuzzy logic controller* (IT2 FLC) (Hagras 2004), *interval type-2 fuzzy expert system*, or *interval type-2 fuzzy model*.

The inputs to the IT2 FLS prior to fuzzification may be certain (e. g., perfect measurements) or uncertain (e. g., noisy measurements). T1 or IT2 FSs can be used to model the latter measurements.

The IT2 FLS works as follows: the crisp inputs are first fuzzified into either type-0 (known as *singleton fuzzification*), type-1 or IT2 FSs, which then activate the inference engine and the rule base to produce output IT2 FSs. These IT2 FSs are then processed by a type-reducer (which combines the output sets and then performs a centroid calculation), leading to an interval T1 FS called the *type-reduced set*. A defuzzifier then defuzzifies the type-reduced set to produce crisp outputs.



Fuzzy Logic, Type-2 and Uncertainty, Fig. 7 Type-2 fuzzy logic system

Rules are the heart of a FLS, and may be provided by experts or can be extracted from numerical data. In either case, rules can be expressed as a collection of IF-THEN statements. A *multi-inputmulti-output* (MIMO) rule base can be considered as a group of *multi-input single-output* (MISO) rule bases; hence, it is only necessary to concentrate on a *MISO* rule base. Consider an IT2 FLS having p inputs $x_1 \in X_1, \dots, x_p \in X_p$ and one output $y \in Y$. We assume there are M rules where the i th rule has the form

$$R^i : \text{IF } x_1 \text{ is } \tilde{F}_1^i \text{ and } \dots \text{ and } x_p \text{ is } \tilde{F}_p^i, \text{ THEN } y \text{ is } \tilde{G}^i \\ i = 1, \dots, M. \quad (17)$$

This rule represents a T2 relation between the input space, $X_1 \times \dots \times X_p$, and the output space, Y , of the IT2 FLS. Associated with the p antecedent IT2 FSs, $\tilde{F}_k^i$, are the IT2 MFs $\mu_{\tilde{F}_k^i}(x_k) (k = 1, \dots, p)$, and associated with the consequent IT2 FS $\tilde{G}^i$ is its IT2 MF $\mu_{\tilde{G}^i}(y)$.

The major result for an interval singleton T2 FLS, i.e. an IT2 FLS in which inputs are modeled as perfect measurements (type-0 FSs, singleton defuzzification) is summarized in the following:

Theorem 2 (Liang and Mendel 2000; Mendel 2001) *In an interval singleton T2 FLS using product or minimum t-norm, for input $\mathbf{x} = \mathbf{x}'$:*

- (a) *The result of the input and antecedent operations, is an IT1 set called the firingset, i.e.,*

$$F^i(\mathbf{x}') = [\underline{f}^i(\mathbf{x}') \tilde{f}^i(\mathbf{x}')] \equiv [\underline{f}^i \tilde{f}^i] \\ = \left[\underline{\mu}_{\tilde{F}_1^i}(x'_1) * \dots * \underline{\mu}_{\tilde{F}_p^i}(x'_p), \right. \\ \left. \bar{\mu}_{\tilde{F}_1^i}(x'_1) * \dots * \bar{\mu}_{\tilde{F}_p^i}(x'_p) \right]. \quad (18)$$

- (b) *The rule R^i fired output consequent set, $\mu_{\tilde{G}^i}(y)$, is the IT2 FS.*

$$\mu_{\tilde{G}^i}(y) = \int_{b^i \in \left[\underline{f}^{is} \underline{\mu}_{\tilde{F}_k^i}(y) \tilde{f}^{is} \bar{\mu}_{\tilde{F}_k^i}(y) \right]} 1/b^i, \quad y \in Y, \quad (19)$$

where $\underline{\mu}_{\tilde{G}^i}(y)$ and $\bar{\mu}_{\tilde{G}^i}(y)$ are the lower and upper membership grades of $\mu_{\tilde{G}^i}(y)$.

- (c) *Suppose that N of the M rules in the IT2 FLS fire, where $N \leq M$, and the combined output fuzzy set, $\mu_{\tilde{G}}(y)$, is obtained by combining the fired output consequent sets by taking the union of the rule R^i fired output consequent sets; then,*

$$\mu_{\tilde{B}}(y) = \int_{b \in \left[\begin{matrix} \left[\begin{matrix} \underline{f}^{1*} & \underline{\mu}_{-1}(y) \\ \underline{f}^{1*} & \underline{\mu}_{-1}(y) \end{matrix} \right] \vee \dots \vee \left[\begin{matrix} \underline{f}^{N*} & \underline{\mu}_{-N}(y) \\ \underline{f}^{N*} & \underline{\mu}_{-N}(y) \end{matrix} \right] \\ \left[\begin{matrix} \underline{f}^{1*} & \underline{\mu}_{-1}(y) \\ \underline{f}^{1*} & \underline{\mu}_{-1}(y) \end{matrix} \right] \vee \dots \vee \left[\begin{matrix} \underline{f}^{N*} & \underline{\mu}_{-N}(y) \\ \underline{f}^{N*} & \underline{\mu}_{-N}(y) \end{matrix} \right] \end{matrix} \right]}^{1/b}, \quad y \in Y. \quad (20)$$

We do not necessarily advocate taking the union of these sets. Part (c) of this theorem merely illustrates the calculations if one chooses to do this. Generalizations of this theorem to an input that is a T1 or an IT2 FS are also given in (Liang and Mendel 2000; Mendel 2001) and (Mendel et al. 2006a).

In Fig. 7, the *type-reduced set* provides an interval of uncertainty for the output of an IT2 FLS, in much the same way that a confidence interval provides an interval of uncertainty for a probabilistic system. The more uncertainties that occur in an IT2 FLS, which translate into more uncertainties about its MFs, the larger will be the type-reduced set, and vice-versa.

Five different type-reduction (TR) methods are described in (Karnik et al. 1999; Mendel 2001). Each is inspired by what is done in a T1 FLS [when the (combined) output of the inference

engine is defuzzified using a variety of defuzzification methods that all do some sort of centroid calculation] and are based on computing the *centroid of an IT2 FS*. Center-of-sets, centroid, center-of-sums, and height type-reduction can all be expressed as

$$\begin{aligned} Y_{\text{TR}}(\mathbf{x}') &= [y_l(\mathbf{x}'), y_r(\mathbf{x}')] \equiv [y_l, y_r] \\ &= \int_{y^1 \in [y_l^1, y_r^1]} \dots \int_{y^M \in [y_l^M, y_r^M]} \int_{f^1 \in [\underline{f}^1, \bar{f}^1]} \\ &\quad \dots \int_{f^M \in [\underline{f}^M, \bar{f}^M]} 1 / \frac{\sum_{i=1}^M f^i y^i}{\sum_{i=1}^M f^i}, \end{aligned} \quad (21)$$

where the multiple integral signs denote the union operation. For a detailed explanation of (21) see (Karnik et al. 1999; Mendel 2001). The most widely used TR is center-of-sets (COS)TR, for which: y_l^i and y_r^i are the left and right end points of the centroid of the consequent of the i th rule [the centroids of all consequent IT2 FSs can be pre-computed using the KM Algorithms (Table 2) and stored for COS TR]; $\underline{f}^i$ and $\bar{f}^i$ are the lower and upper firing degrees of the i th rule, computed using (18); and M is the number of fired rules. For

Fuzzy Logic, Type-2 and Uncertainty, Table 2 KM Algorithms for computing the centroid end-points of an IT2 FS, $\tilde{A}$, and their properties (Karnik and Mendel 2001; Mendel 2001; Mendel and Liu 2007). Note that $x_1 \leq x_2 \leq \dots \leq x_N$

Step	KM Algorithm for c_l $c_l = \min_{\forall \theta_i \in [\underline{\mu}_{-1}(x_i), \bar{\mu}_{-1}(x_i)]} \left(\sum_{i=1}^N x_i \theta_i / \sum_{i=1}^N \theta_i \right)$	KM Algorithm for c_r $c_r = \max_{\forall \theta_i \in [\underline{\mu}_{-1}(x_i), \bar{\mu}_{-1}(x_i)]} \left(\sum_{i=1}^N x_i \theta_i / \sum_{i=1}^N \theta_i \right)$
1	Initialize θ_i by setting $\theta_i = \left[\frac{\underline{\mu}_{-1}(x_i) + \bar{\mu}_{-1}(x_i)}{2} \right] / 2$, $i = 1, \dots, N$ (or $\theta_i = \underline{\mu}_{-1}(x_i)$, $i \leq \lfloor (n+1)/2 \rfloor$ and $\theta_i = \bar{\mu}_{-1}(x_i)$, $i > \lfloor (n+1)/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the first integer equal to or smaller than $\cdot$), and then compute $c' = c(\theta_1, \dots, \theta_N) = \sum_{i=1}^N x_i \theta_i / \sum_{i=1}^N \theta_i$	
2	Find k ($1 \leq k \leq N-1$) such that $x_k \leq c' \leq x_{k+1}$	
3	Set $\theta_i = \bar{\mu}_{-1}(x_i)$ when $i \leq k$, and $\theta_i = \underline{\mu}_{-1}(x_i)$ when $i \geq k+1$, and then compute $c_l(k) \equiv$ $\frac{\sum_{i=1}^k x_i \bar{\mu}_{-1}(x_i) + \sum_{i=k+1}^N x_i \underline{\mu}_{-1}(x_i)}{\sum_{i=1}^k \bar{\mu}_{-1}(x_i) + \sum_{i=k+1}^N \underline{\mu}_{-1}(x_i)}$	Set $\theta_i = \underline{\mu}_{-1}(x_i)$ when $i \leq k$, and $\theta_i = \bar{\mu}_{-1}(x_i)$ when $i \geq k+1$, and then compute $c_r(k) \equiv$ $\frac{\sum_{i=1}^k x_i \underline{\mu}_{-1}(x_i) + \sum_{i=k+1}^N x_i \bar{\mu}_{-1}(x_i)}{\sum_{i=1}^k \underline{\mu}_{-1}(x_i) + \sum_{i=k+1}^N \bar{\mu}_{-1}(x_i)}$
4	Check if $c_l(k) = c'$. If yes, stop and set $c_l(k) = c_l$ and call $k k_L$. If no, go to step 5	Check if $c_r(k) = c'$. If yes, stop and set $c_r(k) = c_r$ and call $k k_R$. If no, go to step 5
5	Set $c' = c_l(k)$ and go to step 2	Set $c' = c_r(k)$ and go to step 2

Properties of the KM Algorithms (Mendel and Liu 2007)

Convergence is monotonic and super-exponentially fast.

other kinds of TR methods, $y_l^i, y_r^i, \underline{f}^i, \bar{f}^i$ and M have different meanings, and are summarized in Table I of (Mendel et al. 2006b).

The defuzzified output of the IT2 FLS is simply the average of y_l and y_r , i. e.

$$y(\mathbf{x}') = [y_l(\mathbf{x}') + y_r(\mathbf{x}')]/2. \quad (22)$$

Because TR is iterative, it may not be possible to use it in a real-time application. Wu and Mendel (Wu and Mendel 2002) introduced a method to approximate the TR set by minimax uncertainty bounds. Doing this avoids the computational overheads associated with TR and, as shown in (Lynch et al. 2005; Wu and Mendel 2002), provides very similar outputs to the IT2 FLSs using TR. These uncertainty bounds are $\underline{y}_l(\mathbf{x}') \leq y_l(\mathbf{x}') \leq \bar{y}_l(\mathbf{x}')$ and $\underline{y}_r(\mathbf{x}') \leq y_r(\mathbf{x}') \leq \bar{y}_r(\mathbf{x}')$, where:

$$\bar{y}_l(\mathbf{x}') = \min \left\{ \frac{\sum_{i=1}^M \underline{f}^i y_l^i}{\sum_{i=1}^M \underline{f}^i}, \frac{\sum_{i=1}^M \bar{f}^i y_l^i}{\sum_{i=1}^M \bar{f}^i} \right\}, \quad (23)$$

$$\bar{y}_r(\mathbf{x}') = \max \left\{ \frac{\sum_{i=1}^M \underline{f}^i y_r^i}{\sum_{i=1}^M \underline{f}^i}, \frac{\sum_{i=1}^M \bar{f}^i y_r^i}{\sum_{i=1}^M \bar{f}^i} \right\}, \quad (24)$$

$$\begin{aligned} \underline{y}_l(\mathbf{x}') &= \bar{y}_l(\mathbf{x}') - \\ &\left[\frac{\sum_{i=1}^M (\bar{f}^i - \underline{f}^i)}{\sum_{i=1}^M \bar{f}^i \sum_{i=1}^M \underline{f}^i} \times \frac{\sum_{i=1}^M \underline{f}^i (y_l^i - y_l^1) \sum_{i=1}^M \bar{f}^i (y_l^M - y_l^i)}{\sum_{i=1}^M \underline{f}^i (y_l^i - y_l^1) + \sum_{i=1}^M \bar{f}^i (y_l^M - y_l^i)} \right], \end{aligned} \quad (25)$$

$$\begin{aligned} \bar{y}_r(\mathbf{x}') &= \underline{y}_r(\mathbf{x}') + \\ &\left[\frac{\sum_{i=1}^M (\bar{f}^i - \underline{f}^i)}{\sum_{i=1}^M \bar{f}^i \sum_{i=1}^M \underline{f}^i} \times \frac{\sum_{i=1}^M \bar{f}^i (y_r^i - y_r^1) \sum_{i=1}^M \underline{f}^i (y_r^M - y_r^i)}{\sum_{i=1}^M \bar{f}^i (y_r^i - y_r^1) + \sum_{i=1}^M \underline{f}^i (y_r^M - y_r^i)} \right]. \end{aligned} \quad (26)$$

Observe that the four bounds in (23)–(26) can be computed without having to perform TR. Wu and Mendel (Wu and Mendel 2002) then approximate the TR set, as $[y_l(\mathbf{x}'), y_r(\mathbf{x}')] \approx$

$\left[\left(\underline{y}_l(\mathbf{x}') + \bar{y}_l(\mathbf{x}') \right) / 2, \left(\underline{y}_r(\mathbf{x}') + \bar{y}_r(\mathbf{x}') \right) / 2 \right]$ and compute the output of the IT2 FLS as

$$\begin{aligned} y(\mathbf{x}') &= \frac{1}{2} \\ &\times \left[\frac{\underline{y}_l(\mathbf{x}') + \bar{y}_l(\mathbf{x}')}{2} + \frac{\underline{y}_r(\mathbf{x}') + \bar{y}_r(\mathbf{x}')}{2} \right] \end{aligned} \quad (27)$$

(instead of as in (22)). So, by using the uncertainty bounds, they obtain both an approximate TR set as well as a defuzzified output.

Wu and Mendel (Wu and Mendel 2002) still use TR during the design of an IT2 FLS. They define a new objective function that trades off some RMSE with not having to perform TR during the real-time operation of the IT2 FLS. The drawback to this approach is that TR is still performed during the design step. Lynch et al. (Lynch et al. 2005, 2006) abandon TR completely where they replace all of the IT2 FLS computations with those in (23)–(26) which gave very similar outputs to the IT2 FLSs using TR. Doing this leads to an IT2 FLS real-time architecture.

Applications

Applications for IT2 FLSs or IT2 FSs are very numerous. Those that have appeared in the literature prior to 2001 can be found in (see pp. 13–14 in (Mendel 2001)), and those that have appeared between 2001 and 2006 can be found in (Mendel 2007a) and (see Table 24.8 in (Mendel 2008)). It is worth mentioning that applications have now appeared for the following general classes of problems: approximation, clustering, communications, control, databases, decision making embedded agents, health care, hidden Markov models, knowledge mining, neural networks, pattern classification, quality control, scheduling, signal and image processing, and spatial query. Control applications, which were the original bread-and-butter ones for T1 FLSs, are now a major focus of attention for IT2 FLSs, and even general T2 FLSs (Hagras 2007). provides three important applications that demonstrate that an IT2 FLS can significantly outperform a T1 FLS. Recently, IT2 FLSs have also been implemented in hardware (Melgarejo and Pena-Reyes 2007); this should make them more attractive and accessible for industrial applications.

Future Directions

There is much work still to be done in type-2 fuzzy research. But, we see that particular areas for fruitful and interesting work will include:

1. Applications. The two broad areas that fuzzy logic is used we can categorize as control and non-control. As discussed there is already some work in using type-2 for control and we expect this to grow. However, the weakness in type-1 fuzzy logic applications is in non-control where we are trying to emulate human expertise. Type-2 fuzzy sets are well placed to help here.
2. Generalized type-2 fuzzy systems. The relatively new approaches for allowing generalized type-2 fuzzy systems is exciting and we expect many researchers to exploit these approaches in interesting applications and provide new theoretical results.
3. Computing with Words. Because words mean different things to different people, we strongly believe that type-2 fuzzy sets must be used to model words when implementing Zadeh's Computing with Words paradigm.

Bibliography

- Bustince H (2000) Indicator of inclusion grade for interval-valued fuzzy sets: application to approximate reasoning based on interval-valued fuzzy sets. *Int J Approx Reason* 23:137–209
- Coupland S (2007) Type-2 fuzzy sets: geometric defuzzification and type-reduction. In: *Proceedings of FOCI*, pp. 622–629
- Coupland S, John RI (2005) Towards more efficient type-2 fuzzy logic systems. In: *Proceedings of Fuzz-IEEE*, pp 236–241
- Coupland S, John RI (2007) Geometric type-1 and type-2 fuzzy logic systems. *IEEE Trans Fuzzy Syst* 15(1): 3–15
- Gorzalczany MB (1987) Decision making in signal transmission problems with interval-valued fuzzy sets. *Fuzzy Sets Syst* 23:191–203
- Gorzalczany MB (1988) Interval-valued fuzzy controller based on verbal model of object. *Fuzzy Sets Syst* 28: 45–53
- Hagras H (2004) A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots. *IEEE Trans Fuzzy Syst* 12:524–539
- Hagras H (2007) Type-2 FLCs: a new generation of fuzzy controllers. *IEEE Comput Intel Mag* 2:30–43
- Karnik NN, Mendel JM (2001) Centroid of a type-2 fuzzy set. *Inf Sci* 132:195–220
- Karnik NN, Mendel JM, Liang Q (1999) Type-2 fuzzy logic systems. *IEEE Trans Fuzzy Syst* 7:643–658
- Liang Q, Mendel JM (2000) Interval type-2 fuzzy logic systems: theory and design. *IEEE Trans Fuzzy Syst* 8: 535–550
- Lynch C, Hagras H, Callaghan V (2005) Embedded type-2 FLC for real-time speed control of marine and traction diesel engines. In: *Proceedings of FUZZ-IEEE*, pp. 347–353
- Lynch C, Hagras H, Callaghan V (2006) Using uncertainty bounds in the design of embedded real-time type-2 neuro-fuzzy speed controller for marine diesel engines. In: *Proceedings of FUZZ-IEEE*, pp 7217–7224
- Melgarejo M, Pena-Reyes CA (2007) Implementing interval type-2 fuzzy processors. *IEEE Comput Intel Mag* 2: 63–71
- Mendel JM (2001) Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice-Hall, Upper Saddle River
- Mendel JM (2007a) Advances in type-2 fuzzy sets and systems. *Inf Sci* 177:84–110
- Mendel JM (2007b) Type-2 fuzzy sets and systems: an overview. *IEEE Comput Intel Mag* 2:20–29
- Mendel JM (2008) On type-2 fuzzy sets as granular models of words. In: Pedrycz W, Skowron A, Kreinovich V (eds) *Granular computing handbook*. Wiley, London
- Mendel JM, John RI (2002) Type-2 fuzzy sets made simple. *IEEE Trans Fuzzy Syst* 10:117–127
- Mendel JM, Liu F (2007) Super-exponential convergence of the Karnik-Mendel algorithms for computing the centroid of an interval type-2 fuzzy set. *IEEE Trans on Fuzzy Syst* 15:309–320
- Mendel JM, Wu H (2007) New results about the centroid of an interval type-2 fuzzy set, including the centroid of a fuzzy granule. *Inf Sci* 177:360–377
- Mendel JM, John RI, Liu F (2006a) Interval type-2 fuzzy logic systems made simple. *IEEE Trans Fuzzy Syst* 14: 808–821
- Mendel JM, Hagras H, John RI (2006b) Standard background material about interval type-2 fuzzy logic systems that can be used by all authors. IEEE Computational Intelligence Society standard. <http://iee.cis.org/standards>
- Wu H, Mendel JM (2002) Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems. *IEEE Trans Fuzzy Syst* 10:622–639
- Wu D, Mendel JM (2007) Uncertainty measures for interval type-2 fuzzy sets. *Inf Sci* 177:5378–5393
- Wu D, Mendel JM (2008) A vector similarity measure for linguistic approximation: interval type-2 and type-1 fuzzy sets, vol 177. *Inf Sci* 178:381–402
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning–1. *Inf Sci* 8: 199–249

Fuzzy Optimization

Weldon A. Lodwick and Elizabeth A. Untiedt
Department of Mathematical Sciences, University
of Colorado Denver, Denver, CO, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Classical Approaches to Fuzzy Optimization
Possibilistic, Interval, Cloud, and Probabilistic
Optimization Utilizing IVP
Future Directions
Bibliography

Glossary

Fuzzy set A set whose membership is characterized by gradualness and uniquely described by a membership function which measures the degree of membership that domain values possess with respect to belonging to the set in question.

Fuzzy set theory The theory of uncertainty associated with sets characterized by gradual membership.

Possibility theory The theory of uncertainty associated with deficiency of information.

Fuzzy number A fuzzy set whose membership function is upper semi-continuous with a unique modal value and whose domain is the set of real numbers.

Possibilistic number A variable described by a possibilistic distribution whose domain is the set of real numbers.

Optimization The mathematical field that studies normative processes.

Definition of the Subject

Fuzzy optimization is normative and as a mathematical model deals with transitional uncertainty

and information deficiency uncertainty. Some literature calls these uncertainties *vagueness* and *ambiguity* respectively. Transitional uncertainty is the domain of *fuzzy set theory* while uncertainty resulting from information deficiency is the domain of *possibility theory*. Suppose one is told by one's employer to go to the airport and meet a tall female visitor at the baggage claim of the airport. The notion of "tall" is transitional uncertainty. As the passengers arrive at the baggage claim, one compares each female passenger to the ascribed characteristic (tall woman) as determined by one's evaluation of what the boss' function is for "tall woman" and all the information one might have regarding "tall women". The resulting function is used to obtain a possibility value for each female that appears at the baggage claim. This uncertainty arises from information deficiency.

An example of fuzzy and possibilistic uncertainty in optimization is in the radiation treatment plan of tumors which is a problem that seeks to deposit a tumorcidal dose to cells that are cancerous while sparing all other cells and at the same time minimizing the total amount of radiation used for the treatment. In this context fuzzy uncertainty arises in cell classification because a cell may be precancerous in which case it is both healthy and cancerous at the same time to some degree. Typically, radiation treatment modelling discretizes a CT-scan of a patient into pixels (voxels in three dimensions) where a pixel is a position in space relative to a fixed coordinate system containing the patient and the radiation machine. The nature of a particular pixel in space may be two (or more) things at once. That is, a particular pixel may be precancerous and thus cancerous and healthy at the same time. Classification into discrete states leaves open transitional states. Continuous states are often intractable. That is, given a discrete classification of cells, what type of cell a pixel models is often transitional and thus fuzzy. Optimization models that incorporate uncertainty of these types are confronted with a wider spectrum of uncertainty

than merely uncertainty due to probability, that is, frequency.

Possibilistic uncertainty arises when there is the specification, “deposit 60 units of radiation at every tumor cell”. The minimal radiation that kills a tumor cell is considered 60 units of radiation by the community of radiation oncologists. The number 60 is derived from mathematical models, research, experience, and expert knowledge. It (60 units) represents the best available information as to a minimum radiation dose that will kill a tumor cell. Of course, if a radiation oncologist were able to attain 59.999 or 60.001 but not 60 units at a tumor cell, this would undoubtedly be satisfactory. The 60 units is “informational” since it is derived from research results and experience whose value as a single mathematical entity, the real number 60, is not precise in fact. It is not just a matter of measurement or probabilistic/statistical uncertainty though both statistics and probability theory may play a part in the determination of the number 60. The number 60 is possibilistic rather than probabilistic (or fuzzy) since the uncertainty associated with the death of a cancer cell given a precise amount of radiation, 60 in this case, as cause/effect (certainly killing a cancerous cell as a result of delivering precisely 60 units of radiation to the cell) is derived from sources beyond frequency analysis. The death of a cancer cell, as a result of radiation, certainly must depend on the type of cancer, its stage of growth, personal genetic characteristics, amount of food in the blood supply at the time of radiation, and so on.

Mathematical models may be considered as being of two types – descriptive (such as simulation) and normative. Optimization models and problems, including fuzzy optimization, are normative in that they impose upon the mathematical system criteria that seek to extract a “best”.

This exposition will clearly identify fuzzy optimization as a distinct optimization approach. While it is not exhaustive, this presentation will focus on the salient features of fuzzy optimization most related to optimization under uncertainty. In particular, fuzzy multiobjective programming, fuzzy stochastic programming, and fuzzy dynamic programming are not discussed (see,

for example Liu 1999, 2002; Luhandjula 2004, 2006; Sahinidis 2004). Since intervals (and real numbers) may be considered as fuzzy sets, interval optimization is not covered separately but considered within the family of fuzzy optimization.

Mathematical analyzes that include the normative must embody the idea of order, measure, and distance. The notion of “best” requires an order and measuring with respect to that order. Professor Lothar Collatz in a lecture titled, “Monotonicity and Optimization”, given to the Mathematics Department at Oregon State University, August 5, 1974, stated, “The idea of *ordering* is more fundamental than *distance* because from ordering we can get distance but not the other way around”. (my notes) The real number system contains within itself the most fundamental mathematical order. Optimal control, stochastic dominance, (Ogryczak and Ruszczyński 1999; Whitmore and Findlay 1978), and mean-variance, (Markowitz 1952; Ogryczak and Ruszczyński 1999; Steinbach 2001), are approaches to order functions. Since fuzzy intervals from their definition (see below) relate themselves to *sets* of real numbers, that is, graphs in $\mathbb{R}^2$ (membership functions), the order of fuzzy intervals will need to be derived from that associated with real-valued functions. Moreover, since fuzzy sets model uncertainty on one hand and amplification and flexibility on the other, it is clear that the idea of order and its derived distance (measure) generated will be flexible, that is, require choices, as we shall see. The choice that needs to be made is dependent on the semantics of the problem to a greater extent than in the deterministic setting.

Mathematical modeling involves simplification, the transformation from reality to symbols. Even after a symbolic representation of a system, the process being modeled may, for example, be nonlinear, but is only tractable if linearized. Moreover, a nonlinear process may be the correct model but its linear counterpart may yield acceptable solutions. In the case of radiation therapy models (see Lodwick et al. 1999, 2001), scatter, which causes nonlinearities, is typically ignored in developing radiation therapy models that determine the angles and intensities for a given

radiation machine and a given patient being treated. Without scatter, the resulting model is a linear model whose results usually produce acceptable treatment plans as long as the breathing of the patient is ignored. On the other hand, obtaining the mathematical model for the location of a cell (in a fixed coordinate system that includes both the patient and the radiation machine) of a patient undergoing radiation therapy, is important to include (while scatter may not be as important in the determination of acceptable angles and intensities). Location of points in the body that are or have been affected by breathing is not a deterministic model. Nor is it a probabilistic model.

The mathematical model development for radiation therapy planning of a particular patient tumor for a particular machine not only is designed to “do the job” (kill all tumor cells while sparing healthy cells) but often adds various “normative” criteria. For example, a radiation oncologist might seek a mathematical model which will, when used, kill the tumor cells, spare the healthy cells, *and* minimize total radiation used to do this. The “minimize total radiation used” is a normative criterion. Of course, there could be many other normative criteria imposed, such as “minimize the probability that healthy cells will become cancerous by the radiation deposited from the radiation treatment”.

The thesis of this presentation is twofold. First, not all uncertainty that occurs and is incorporated in optimization models can be described by the frequency with which its parameters take on various values (probability). Fuzzy and possibility theory are necessary components in some cases (such as in radiation of tumor models). Secondly, optimization under uncertainty models at their most general symbolic level try to capture the true complexity of the systems being modeled so that it is crucial to model transitional processes as fuzzy sets, frequencies as probabilistic distributions, and information deficiencies as possibility distributions. Once this is done, a further simplification may be (usually is) involved, but at the highest symbolic level, it is crucial to be faithful to the nature of the uncertainty, since this will ultimately determine the correct semantics and

correct simplifying assumptions. For example, a normative probability model may be translated into a stochastic recourse model which is transformed into a real-valued nonlinear programming problem. Knowing that the underlying process is probabilistic means that the input data must be faithful to the laws of probability even though in the end, the model is a real-valued nonlinear programming model. The solution semantics in this case are those arising from probability and thus frequency based. A fuzzy linear optimization model is often translated into a real-valued linear programming model. Knowing that the underlying processes being modeled are fuzzy means that the nature of both the input data and the semantic interpretation of the output solution are based on the laws of fuzzy set theory and not of probability. Fuzzy optimization models are frequently solved as a real-valued linear or nonlinear programming problems just as in the case of stochastic optimization. This process, insuring that at the highest symbolic level of mathematical modeling the correct uncertainty is used and only then transforming it into a real-value model, is analogous to first modeling a nonlinear system as a nonlinear system of relations and then linearizing it. It is crucial to create the nonlinear model first (at least in principle) and then to linearize it so that the approximation that is being used is explicit. When one is clear about these steps, one is able to take into account the correct associated approximation errors and underlying assumptions which enable the appropriate interpretation of the solution output.

Likewise, normative models of fuzzy processes require, at the level of mathematical abstraction, fuzzy modeling *before* approximation and transformation to an equivalent real-valued model because the fuzzy model is most faithful to the underlying uncertainty and thus is able to adhere to the basic assumptions associated with its uncertainty type. Moreover, if the starting point is a fuzzy model, when approximations and translations are made, the sources and magnitudes of associated errors are explicit. Lastly, the solution semantics are determined by the context of the input uncertainty. The modelling of normative processes that contain transitional and/or

information deficiency uncertainty should utilize fuzzy and/or possibilistic optimization. The symbolic representation faithfully executed according to the associated axioms governing the uncertainty type and its semantics is a necessary first step. That is, mathematics (or any science) tries to bare all of its underlying assumptions. Since mathematical models objectify relationships occurring in reality via symbols, the nature of uncertainty must first be made explicit and then approximated, not the other way around.

Fuzzy optimization, which for this exposition encompasses models affected by both fuzzy and possibilistic uncertainty, is one of the newest optimization fields. Its place is along side stochastic optimization within the field of optimization under uncertainty. Fuzzy optimization began in 1970 with the publication of the seminal Bellman and Zadeh paper (1970). It took three years before the next fuzzy optimization article was published in 1973 by H. Tanaka, T. Okuda, and K. Asai, (Asai and Tanaka 1973; Tanaka et al. 1973), (with the full version (Tanaka et al. 1974)). These researchers seem to have been the first to realize the importance of alpha-levels in the mathematical analysis of fuzzy sets in general and fuzzy optimization in particular. The Tanaka, Okuda, Asai article operationalized the theoretical approach developed by Bellman and Zadeh. Independently, in 1974, H.-J. Zimmermann presented a paper at the *ORSA/TIMS* conference in Puerto Rico (Zimmermann 1974) (with the full version (Zimmermann 1976)) that not only operationalized the Bellman and Zadeh approach, but greatly simplified and clarified fuzzy optimization, so much so that Zimmermann's approach is a standard to this day. In this same period, the book by C.V. Negoita and D.A. Ralescu (1975) contained a description of fuzzy optimization. C.V. Negoita and M. Sularia published in 1976 a set containment approach to fuzzy optimization (Negoita and Sularia 1976). From this beginning, fuzzy optimization has become a field of study in its own right with a journal devoted to the subject, *Fuzzy Optimization and Decision Making*, whose first issue came out in February of 2002. Moreover, there have been special sessions devoted solely to fuzzy/possibilistic optimization at the

international fuzzy society meetings (*IFSA05*, July 2005, Beijing, China and *IFSA07* June 2007, Cancun, Mexico). There have been two special editions of the journal *Fuzzy Sets and Systems* dealing with fuzzy/possibilistic optimization, the latest being (Lodwick and Inuiguchi 2007). Two books with edited articles have appeared – (Delgado et al. 1994; Kacprzyk and Orlovski 1987a) and there are at least six authored books devoted to fuzzy optimization – (Bector and Chandra 2005; Lai and Hwang 1992; Liu 1999, 2002; Ramik and Vlach 2002b; Sousa and Kaymak 2002).

Thirty-five years of fuzzy optimization research has yielded a wide-ranging set of applications. It is beyond the scope of this exposition to cover applications, but, the interested reader may wish to consult the following set of references: Chaps. 6 and 8 in (Delgado et al. 1994), Chap. III in (Kacprzyk and Orlovski 1987a), and (Bector and Chandra 2005; Ertuğrul and Tuş 2007; Gladish et al. 2005; Hsu and Wang 2001; Inuiguchi and Ramik 2000; Inuiguchi et al. 1994; Joubert et al. 2007; Lodwick 2004; Lodwick and Bachman 2005; Lodwick and Inuiguchi 2007; Lodwick and Jamison 2003a; Lodwick et al. 1999, 2001; Riverol and Pilipovik 2007; Riverol et al. 2007; Rommelfanger 1996; Sahinidis 2004; Tang et al. 2001; Untiedt 2006; Vasant et al. 2007; Wang and Liang 2005; Wang and Zhu 2002).

The general fuzzy optimization model considered in this presentation is to find a fuzzy optimum of a fuzzy objective function subject to a fuzzy constraint,

$$\widetilde{\text{opt}}_x \tilde{z} = f(\tilde{c}, x) \quad (1)$$

$$x \in \tilde{X} \quad (2)$$

where the tilde, $\sim$, represents fuzzy and/or possibilistic entities or relationships which is made clear by the context and assumptions of the problem. When fuzzy uncertainty needs to be distinguished from possibilistic uncertainty, the tilde, $\sim$, will denote fuzzy uncertainty and the circumflex, $\hat{\sim}$, will denote possibilistic uncertainty. For (1), $\tilde{c}$ is considered to be a known vector of

uncertainty parameters characterized by a fuzzy membership function or a possibilistic distribution. The variables (unknown quantities whose values are to be determined by the model) which are denoted here by x , are often called the “decision variables” because in optimization models, it is the value of x that is being computed. For example, x may denote a quantity to be produced by a manufacturing process, or the amount to be transported from a production point to a consuming point, or the intensity of radiation for the angle represented by the variable, and so on.

Introduction

It is assumed that the reader has a basic knowledge of fuzzy set theory at the level of (Klir and Yuan 1995) though the basics that are needed for this exposition are set forth. Much of the introductory exposition can also be found in (Lodwick 2007).

Basics of Fuzzy Set Theory

Fuzzy set and possibility theory were defined and developed by L. Zadeh beginning with (Zadeh 1965) and subsequently (Zadeh 1968, 1975). As is now well-known, the idea was to mathematize and develop analytical tools to solve problems whose uncertainty went beyond probability theory. Classical mathematical sets, for example a set A , have the property that either an element $x \in A$ or $x \notin A$ but not both. There are no other possibilities for classical sets which are also called *crisp* sets. An interval is a classical set. L. Zadeh’s idea was to relax this “all or nothing” membership in a set to allow for grades of belonging to a set. When grades of belonging are used, a fuzzy set ensues. To each fuzzy set, $\tilde{A}$, L. Zadeh associated a real-valued *membership function* $\mu_{\tilde{A}}(x)$, which takes x in the domain of interest, the universe Ω , to a value in the interval $[0, 1]$. The membership function $\mu_{\tilde{A}}(x)$ quantifies the degree to which x belongs to $\tilde{A}$ where a value of zero means that x certainly does not belong to $\tilde{A}$ and a value of one means that x certainly belongs to $\tilde{A}$.

$$\mu_{\tilde{A}}(x): \mathbb{R} \rightarrow [0, 1].$$

Another way of looking at a fuzzy set is as a set in $\mathbb{R}^2$ as follows.

Definition 1 A **fuzzy set** $\tilde{A}$, as a crisp set in $\mathbb{R}^2$, is the set of ordered pairs

$$\tilde{A} = \left\{ \left(x, \mu_{\tilde{A}}(x) \right) \right\} \subseteq \{(-\infty, \infty) \times [0, 1]\}. \quad (3)$$

The α – **cut** of a fuzzy set is the set

$$\tilde{A}_\alpha = \left\{ x \mid \mu_{\tilde{A}}(x) \geq \alpha \right\}.$$

Definition 2 A **modal value** of a membership function is a domain value at which the membership function is one. A fuzzy set with at least one modal value is called **normal**. *The support of a membership function is the closure of* $\left\{ x \mid \mu_{\tilde{A}}(x) > 0 \right\}$.

Definition 3 (see Fortin et al. 2008) A **fuzzy interval**, $\tilde{M}$, defined by its membership function $\mu_{\tilde{M}}(\cdot)$, is a fuzzy continuous subset of the real line such that, if $x, y, z \in \mathbb{R}$, $z \in [x, y]$, then

$$\mu_{\tilde{M}}(z) \geq \min \left\{ \mu_{\tilde{M}}(x), \mu_{\tilde{M}}(y) \right\}.$$

Like a fuzzy set, a fuzzy interval M is said to be **normal** if $\exists x \in \mathbb{R}$ such that $\mu_{\tilde{M}}(x) = 1$. The set $\left\{ x \mid \mu_{\tilde{M}}(x) = 1 \right\}$ is called the **core** (of the fuzzy interval).

Definition 4 A **fuzzy number** is a fuzzy interval with a unique modal value, that is, the core is a singleton.

For all that follows, fuzzy intervals will be assumed to be normal fuzzy intervals with upper semi-continuous membership functions. This means that the α – *cut* of a fuzzy interval, M_α , is a closed interval. Let $M_1 = \left\{ x \mid \mu_{\tilde{M}}(x) = 1 \right\} = [m_1^-, m_1^+]$, be the core of a fuzzy interval $\tilde{M}$ and the open support $M_0 = \{x \mid \mu_{\tilde{M}}(x) > 0\} =$

(m_0^-, m_0^+) . For a fuzzy interval M , $\mu_M(x)$, is non-decreasing for $x \in (-\infty, m_1^-]$ and $\mu_M(x)$ is non-increasing for $x \in [m_1^+, \infty)$.

The fact that we have closed intervals at each α – cut means that fuzzy arithmetic can be defined by interval arithmetic (see Moore 1979) on each α – cut. Unbounded intervals can be handled by extended interval arithmetic. In fact, when dealing with fuzzy intervals, the operations and analysis can be considered as interval operations and analysis on α – cuts (Lodwick 2007). We define the most common types of fuzzy intervals next.

Definition 5 A **triangular fuzzy number**, $\tilde{A} = (\alpha, \beta, \gamma)$, has a membership function μ_A centered at a value α , with a support (β, γ) such that

$$\mu_A(x) = \begin{cases} 1, & x = \alpha, \\ 0, & x \leq \alpha - \beta, \\ 0, & x \geq \alpha + \gamma, \\ 1 - \frac{x - \alpha}{\beta}, & x \in (\alpha - \beta, \alpha) \\ 1 - \frac{\gamma - x}{\gamma}, & x \in (\alpha, \alpha + \gamma). \end{cases}$$

Definition 6 A **symmetric triangular fuzzy number**, $\tilde{A} = (\alpha, \beta)$, has a membership function μ_A centered at a value α , with a spread β such that

$$\mu_A(x) = \begin{cases} 1, & x = \alpha, \\ 0, & x \leq \alpha - \beta, \\ 0, & x \geq \alpha + \beta, \\ 1 - \frac{|x - \alpha|}{\beta}, & x \in (\alpha - \beta, \alpha + \beta). \end{cases}$$

Definition 7 A **trapezoidal fuzzy interval**, $\tilde{A} = (\alpha, \beta, \gamma, \delta)$, has a membership function μ_A with a core (α, β) and a support (γ, δ) such that

$$\mu_A(x) = \begin{cases} 1, & x \in [\alpha, \beta], \\ 0, & x \notin (\gamma, \delta), \\ 1 - \frac{\alpha - x}{\alpha - \gamma}, & x \in (\gamma, \alpha), \\ 1 - \frac{x - \beta}{\delta - \beta}, & x \in (\beta, \delta). \end{cases}$$

Definition 8 A **L-R fuzzy interval**, $\tilde{A} = (\alpha^R, \alpha^L, \beta^R, \beta^L)_{LR}$ has a membership function μ_A and reference functions L and $R: [0, \inf] \rightarrow [0, 1]$ and $R: [0, \inf] \rightarrow [0, 1]$ are upper semi-continuous and strictly decreasing in the range $(0, 1]$, and μ_A is defined as follows:

$$\mu_A(x) = \begin{cases} 1, & x \in [\alpha^L, \alpha^R], \\ 0, & x \notin (\alpha^L - \beta^L, \alpha^R + \beta^R), \\ L\left(\frac{\alpha^L - x}{\beta^L}\right), & x \in [\alpha^L - \beta^L, \alpha^L], \\ R\left(\frac{x - \alpha^R}{\beta^R}\right), & x \in [\alpha^R, \alpha^R + \beta^R]. \end{cases}$$

The following relations are applicable to fuzzy intervals:

Definition 9 (Equality) Two fuzzy sets, $\tilde{A}$ and $\tilde{B}$ are said to be equal if and only if $\mu_A(x) = \mu_B(x)$ for all $x \in X$.

This definition, given by Bellman and Zadeh (Bellman and Zadeh 1970), is generally accepted. We explore broader interpretations of $\tilde{A} = \tilde{B}$ in section “Fuzzy Relations”.

Definition 10 (Containment) A fuzzy set $\tilde{A}$ is said to be a subset of fuzzy set $\tilde{B}$ if and only if $\mu_A(x) \leq \mu_B(x)$ for all $x \in X$.

Definition 11 (Intersection) The *intersection* of $\tilde{A}$ and $\tilde{B}$ is defined as the largest fuzzy set contained in both $\tilde{A}$ and $\tilde{B}$. The membership function of $\tilde{A} \wedge \tilde{B}$ is given by

$$\mu_{A \wedge B}(x) = \min(\mu_A(x), \mu_B(x)), \quad x \in X.$$

Minimum is only one of a continuum of operators that define intersection, but this discussion is beyond the scope of this paper. The *t-norms* (see Klir and Yuan 1995) define a family of intersection operators.

Definition 12 (Relation) A *fuzzy relation* R in the product space $X \times Y$ is a fuzzy set

characterized by a membership function μ_R which associates with each ordered pair a grade of membership $\mu_R(x, y)$ in $\mathbb{R}$.

Definition 13 (Non-Interactivity) Consider a fuzzy number $\tilde{C}$ where $C \subseteq \mathbb{R}^2$ which is a direct product of two fuzzy numbers $\tilde{A} \in \mathbb{R}$ and $\tilde{B} \in \mathbb{R}$ such that for all $(x, y) \in \mathbb{R}^2$,

$$\mu_C(x, y) = \mu_A(x) \wedge \mu_B(y). \quad (4)$$

If condition (4) holds for $\tilde{A}$ and $\tilde{B}$, then $\tilde{A}$ and $\tilde{B}$ are said to be non-interactive. Semantically, two numbers are non-interactive if they can be assigned values independently of each other (Dubois 1987). Every fuzzy and possibilistic model examined in this paper operates on the assumption of non-interactivity of all uncertain entities.

A different and more recent approach to fuzzy entities is possible. Instead of considering a fuzzy interval as a specialized fuzzy set over the set of real numbers, $\mathbb{R}$, Dubois and Prade (see Dubois and Prade 2005) and Fortin, Dubois, and Fargier (see Fortin et al. 2008) introduce the concept of *gradual numbers* to revise the theory of fuzzy intervals so that a (real-valued) fuzzy interval is to a (real-valued) interval what a fuzzy set is to a (classical) set. This restores the algebraic structure of the real numbers to fuzzy arithmetic. (In particular, the usual fuzzy arithmetic which uses interval arithmetic on α – cuts (Kaufmann and Gupta 1985) lacks an additive identity, a multiplicative identity, and the distributive law, all properties of algebra of real numbers.) An earlier approach to fuzzy arithmetic which also restores the algebraic structure of real numbers to fuzzy arithmetic is constraint interval arithmetic on α – cuts (Lodwick 1999, 2007).

Gradual numbers were developed by (Dubois and Prade 2005; Fortin et al. 2008) not only to restore a richer algebraic structure to fuzzy intervals, but to put fuzzy intervals on a firmer foundation. The theory of gradual numbers is one in which the relationship between real-valued intervals and fuzzy intervals is made quite apparent, clear, and compelling. Special (extreme) gradual

numbers serve as endpoints of fuzzy intervals in the same way that real numbers serve as the endpoints of real intervals. The theory of gradual numbers also allows one to deal with the concept of a fuzzy element (of a fuzzy set) in a theoretically sound and meaningful way. The special gradual numbers associated with the endpoints of a fuzzy interval may be thought of as being composed of two parts. The left gradual number endpoint of a fuzzy interval $\tilde{A}$ is the inverse of the membership function μ_A^- restricted to $(-\infty, m_1^-]$. That is, it is the inverse of

$$\mu_A^-(x) = \mu_A(x), \quad x \in (-\infty, m_1^-], \quad (5)$$

where $[m_1^-, m_1^+]$ is the core as before, which is non-empty by definition. The second part, the right gradual number endpoint of $\tilde{A}$, is the inverse of the membership function restricted to $[m_1^+, \infty)$. That is, it is the inverse of

$$\mu_A^+(x) = \mu_A(x), \quad x \in [m_1^+, \infty). \quad (6)$$

Definition 14 (see Fortin et al. 2008) A **gradual number** $\tilde{r}$ is defined by an assignment $A_{\tilde{r}}$ from $(0, 1]$ to $\mathbb{R}$.

Note that for a fuzzy interval, $\tilde{A}$, the functions that define the endpoints, $(\mu_A^-)^{-1}(x)$ (5), and $(\mu_A^+)^{-1}(x)$ (6) are special cases of gradual numbers. Briefly, a gradual number in fuzzy interval $\tilde{A}$ is simply the inverse relation of any unique assignment $r(x): x \in \tilde{A}$ such that $\mu_A^-(x) \leq r(x) \leq \mu_A^+(x)$. For the purpose of optimization, continuous strictly-monotonic assignment functions are of interest. Fuzzy intervals may be defined from the point of view of gradual numbers and in this context, we have the following.

Definition 15 (see Fortin et al. 2008) Using the notion of gradual number, a **fuzzy interval** M is an ordered pair of gradual numbers $(\tilde{m}^-, \tilde{m}^+)$ where $\tilde{m}^-$ is called the fuzzy lower bound or **left**

profile and $\tilde{m}^+$ is called the fuzzy upper bound or **right profile**.

To ensure that the left and right profiles adhere to what has been previously defined as a *fuzzy interval*, several properties of $\tilde{m}^-$ and $\tilde{m}^+$ must hold. In particular (see Fortin et al. 2008):

1. The domains of the assignment functions, $\tilde{A}_m^-$ and $\tilde{A}_m^+$, must be in $(0, 1]$.
2. $\tilde{A}_m^-$ must be increasing and $\tilde{A}_m^+$ must be decreasing.
3. $\tilde{m}^-$ and $\tilde{m}^+$ must be such that $\tilde{A}_m^- \leq \tilde{A}_m^+$.

Remark 16 Fuzzy intervals with properties 1–3 above possess well-defined inverses which are functions. Note that the endpoints of a crisp interval $[a, b]$ have constant assignments, that is, $\tilde{A}_m^-(\alpha) = a$ and $\tilde{A}_m^+(\alpha) = b, 0 < \alpha \leq 1$. These are the left and right profiles of the fuzzy interval membership function of a real-valued interval,

$$\mu_{[a,b]}(x) = \begin{cases} 1 & \text{for } x \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

Since it is constant, it has no fuzziness in it and is simply an interval, not a fuzzy interval. An interval that possesses fuzziness has a non-decreasing, non-constant left profile and a non-increasing, non-constant right profile, but a crisp interval has no fuzziness in the left/right profiles (they are horizontal line segments as gradual numbers). That is, the left and right membership function segments are vertical line segments (no fuzziness), whose inverses are horizontal line segments. Each gradual number that is a horizontal line segment, $y = f(x) = a, 0 \leq x \leq 1$, represents the real number (written as a fuzzy set),

$$\mu_a(x) = \begin{cases} 1 & \text{for } x = a \\ 0 & \text{otherwise.} \end{cases}$$

The application of gradual numbers to optimization is just beginning. Two of these are (Kasperski and Zeilinski 2007; Untiedt 2007b).

Extension Principles

Fuzzy extension principles show how to transform real-valued functions into functions of fuzzy sets on one hand and how to compute fuzzy algebraic or fuzzy transcendental expressions on the other. The meaning of fuzzy arithmetic depends directly on the extension principle since arithmetic operations are (continuous) functions over the reals, assuming division by zero is not allowed, and over the extended real numbers, when division by zero is allowed. The fuzzy arithmetic that results from Zadeh's extension principle (Zadeh 1965) and its relationship to interval analysis has an extensive recent development (see Lodwick 2004, 2007; Lodwick and Jamison 2003a). Moreover, there is an intimate interrelationship between the extension principle being used and the analysis that ensues. Since *t-norms* and *t-conorms* capture the way trade-offs among decisions are made in constrained optimization (see Kaymak and Sousa 2003), the way one extends union and intersection via *t-norms* and *t-conorms* will determine the constraint set.

The extension principle within the context of fuzzy set theory was first proposed, developed, and defined in (Zadeh 1965, 1975).

Definition 17 (Extension Principle of L. Zadeh 1965, 1975) Given a real-valued function $f: X \rightarrow Y$, the function over fuzzy sets $f: S(X) \rightarrow S(Y)$, where $S(X)$ (respectively $S(Y)$) is the set of all fuzzy sets of X (respectively Y) is given by

$$\mu_{f(A)}(y) = \sup \left\{ \mu_A(x) \mid y = f(x) \right\} \quad (7)$$

for all fuzzy subsets A of $S(X)$. In particular, if $(X_1, \dots, X_n)$ is a vector of fuzzy intervals, and $f(x_1, \dots, x_n)$ a real-valued function, then

$$\begin{aligned} \mu_{f(X_1, \dots, X_n)} &= \sup_{(x_1, \dots, x_n)} \min_{1 \leq i \leq n} \{ \mu_{X_i}(x_i) \} \mid z \\ &= f(x_1, \dots, x_n). \end{aligned} \quad (8)$$

The extension principle is used to define functions of fuzzy sets by taking a function over the real numbers and extending it to a fuzzy-set-

valued function with fuzzy set arguments in place of the real numbers. If $X \subset \mathbb{R}$, the set-valued function $F: A \subseteq \mathbf{P}(X) \rightarrow \mathbf{P}(Y)$, where $\mathbf{P}(X)$ is the power-set of X and $\mathbf{P}(Y)$ is the power-set of Y , is an *extension function* of real-valued function $f: X \rightarrow Y \subset \mathbb{R}$ if:

$$F(A) = \{f(x) | x \in A\}, \quad (9)$$

and

$$\left[\inf_{x \in A} f(x), \sup_{x \in A} f(x) \right] \subseteq F(A). \quad (10)$$

This latter condition (10) needs to be imposed when these set-valued extension functions are approximated on a computer so that even computationally, it is always true that when a set is “retracted” to a point,

$$F(\{x\}) = f(x) \quad \forall x \in X$$

where F is the (set-valued) extension of (the real-valued) f . Now, if we have a set of fuzzy subsets $\tilde{A}$ of X , to obtain a resulting fuzzy set within the set of all fuzzy subsets of Y under the mapping f , the extension function of f over fuzzy sets is denoted $\tilde{F}$ and the membership function (recall that a fuzzy set is uniquely defined by its membership function) $\mu_{\tilde{F}(\tilde{A})}$ is defined by:

$$\mu_{\tilde{F}(\tilde{A})}(y) = \sup \left\{ \mu_{\tilde{A}}(x) | y = f(x), x \in X, y \in Y \right\}$$

The definition (8) of the extension principle has led to fuzzy arithmetic. Moreover, it is one of the main mechanisms used for fuzzy (interval) analysis. Various researchers have dealt with the issue of the extension principle and amplified its applicability. H. Nguyen (Nguyen 1978) pointed out, in his 1978 paper, that a fuzzy set needs to be defined to be what Dubois and Prade later called a fuzzy interval (see Dubois and Prade 2005; Fortin et al. 2008) in order that

$$\left[f(\tilde{A}, \tilde{B}) \right]_{\alpha} = f(A_{\alpha}, B_{\alpha})$$

where the function f is assumed to be continuous. In particular A_{α} and B_{α} need to be compact (that is closed/bounded intervals) for each α – cuts. Thus, H. Nguyen defined a fuzzy number as one whose membership function is upper semi-continuous and for which the closure of the support is compact. In this case, the α – cuts generated are closed and bounded (compact) sets, that is, real-valued intervals. This is a well-known result in real analysis. That is, when f is continuous, the decomposition by α – cuts can be used to compute $f(\tilde{X}_1, \dots, \tilde{X}_n)$ via interval analysis by (Nguyen 1978) as

$$\left[f(\tilde{X}_1, \dots, \tilde{X}_n) \right]_{\alpha} = f([X_1]_{\alpha}, \dots, [X_n]_{\alpha}).$$

It should be noted that the gradual number representation of fuzzy intervals allows use of the extension principle without resorting to α – cuts.

R. Yager (Yager 1986) pointed out that by looking at functions as graphs (in the Euclidean plane), the extension principle could be extended to include all graphs thus allowing for analysis of what he calls “non-deterministic” mappings, that is, graphs which are not functions. Now, “non-determinism” as is used by Yager can be considered as point-to-set mappings. Thus, Yager implicitly restores the extension principle to a more general setting of point-to-set mappings.

J. Ramik (1986) points out that we can restore L. Zadeh’s extension principle to its most general setting of set-to-set mappings explicitly. In fact, a fuzzy mapping is indeed a set-to-set mapping. He defines the image of a fuzzy set-to-set mapping as being the set of α ’s generated by the function on the α – cuts of the domain.

Lastly, T.Y. Lin’s paper (2005) is concerned with determining the function space in which the fuzzy set generated by the extension principle “lives”. That is, to what space does the range of the fuzzy function belong? The extension principle generates a resultant membership function in the range space. Suppose one is interested in stable controls, then one way to extend is to generate resultant (range-space) membership functions that are continuous. The definition of continuous

function states that small perturbations in the input, that is, domain, cause small perturbations in the output, that is, range, which is one way to view the definition of stability. T.Y. Lin develops conditions that are necessary in order that the range membership function has some desired characteristics (such as continuity or smoothness).

The essential purpose of fuzzy extension principles is to define functions over fuzzy sets so that the resulting range maintains various properties of interest specific to both the function and its fuzzy set input. In optimization, this is crucial in computing the output of objective and constraint functions (1) and (2). The theory and computational methods associated with distribution arithmetic including fuzzy and possibilistic arithmetic is discussed in detail in (Lodwick 2007).

Basics of Possibility Theory

Possibilistic distributions (of fuzzy intervals or sets) encapsulate the most knowledgeable estimate of the possible values of an entity given the available information. This theory was articulated in (Zadeh 1978). Fuzzy membership function values (of fuzzy intervals or sets) describe the degree to which an entity is that value. Note that if the possibility distribution at x is one, this signifies that the best evidence available indicates x is indeed the entity that the distribution describes. Possibility distributions constructed from first principles require nested sets (see, for example Jamison and Lodwick 2002) and normalization. Possibility distributions are normalized since their semantics are tied to existent entities. Since the entity exists, it is always possible for at least one x . For example, if one is hiking from an elevation of 2000 m to an elevation of 3000 m, one must traverse the 2500 m isoline (at least once). That there is a spot at which this occurs is not in question. The location of the spot is and will be dependent on the information at hand (accuracy of the maps, possibly of a global positioning system, compass, altimeter, expert knowledge). Nevertheless, one knows with certainty in this case that the 2500 m isoline is traversed but not where it is traversed.

On the other hand, normalization is not required of fuzzy membership functions. Thus,

not all fuzzy sets can give rise to possibility distributions.

Possibility theory may be derived in at least one of the following ways:

1. Via normalized fuzzy sets (see Zadeh 1978),
2. Axiomatically from fuzzy measures g that satisfy $g(A \cup B) = \max \{g(A), g(B)\}$ (see Dubois and Prade 1988; Klir and Yuan 1995 for example),
3. Via belief functions of Dempster–Shafer theory whose focal elements are normalized and nested (see Klir and Yuan 1995),
4. By construction (via nested sets with normalization, for example nested α – level sets, see Dubois and Prade 1988; Jamison and Lodwick 2002; Klir and Yuan 1995).

The most general derivation of possibility theory (method 2 above) sets up an order among variables with respect to their being an entity. The *magnitudes* associated with this ordering have no significance other than an indication of order. Thus, if $possibility_A(x) = 0.75$ and $possibility_A(y) = 0.25$ all that can be said is that x more likely to be the entity A than y . One cannot conclude that x is three times more likely to be A than y is. This means that for optimization problems, if the possibility distributions were constructed using the most general assumptions, then comparisons among several distributions is problematic. In particular, setting the possibility level to be greater than or equal to a certain fixed value, say $0 \leq \alpha \leq 1$, does not have the same meaning as setting a probability level or a membership function to be at least α . In the former case the α has no inherently meaning (other than if one has a $\beta > \alpha$ one prefers the decision that generated β to that which generated α) whereas in the former, the value of α is meaningful. Because of this, optimization methods must assume that the possibility distributions are constructed according to probability based possibility (see Jamison and Lodwick 2002) since the possibilities that are so constructed do have meaningful distribution value levels. That is, if the possibility level is α , the value of α has a quantitative (not just order) meaning in relation to all values in $[0, 1]$.

There is a companion set-valued function to possibility called *necessity*, when the measure of the underlying space is finite. The necessity set-valued function is defined by

$$\text{necessity}(\tilde{A}) = 1 - \text{possibility}(\tilde{A}^C),$$

where $\tilde{A}^C$ denotes the complement of the fuzzy set $\tilde{A}$. Semantically, the necessity of an event measures the impossibility of the opposite event.

Semantics of Fuzzy Sets and Possibility Distributions in Fuzzy Optimization

This study restricts uncertainty to parameters (input data) whose fuzzy membership functions or possibilistic distributions are over intervals of real numbers, that is, fuzzy or possibilistic intervals. That semantics is crucial in the context of fuzzy sets was known early in the development of fuzzy/possibility theory (Dubois and Prade 1980a) and subsequently elaborated (Dubois and Prade 1986; Dubois et al. 1997).

Fuzzy optimization is distinguished from possibilistic optimization by both semantics and optimization procedures. As will be seen below, fuzzy optimization optimizes over sets of real numbers while possibilistic optimization optimizes over sets of (possibility) distributions. In addition, optimization procedures are influenced by the fact that fuzzy and possibilistic distributions have different development when they are derived from first principles.

The semantic distinctions between fuzzy and possibilistic optimization can be found in (Inuiguchi 1992; Inuiguchi et al. 1992a; Inuiguchi et al. 1994) where it is noted that, for optimization models, *ambiguity* in the coefficients of the model leads to possibility optimization, while *vagueness* in the decision maker's preference is modeled by fuzzy optimization. When this vagueness represents a willingness on the part of the decision maker to relax his or her requirements in order to attain better results, this type of fuzzy optimization is sometimes called *flexible programming*. It has also been said that, in the context of optimization, possibilistic uncertainty is information-based uncertainty and fuzzy uncertainty is

preference-based uncertainty (Lai and Hwang 1992). Another point of view on the semantic distinction between fuzzy and possibilistic uncertainty in optimization is evinced in (Inuiguchi et al. 1992a).

The membership grade of a fuzzy goal (fuzzy constraint) represents the *degree of satisfaction*, whereas that of a possibility distribution represents the *degree of occurrence*.

The semantics of an optimization problem are also influenced by where in the optimization problem the uncertainty occurs. Suppose an optimization problem is known to have fuzzy uncertainty in the inequality constraints. In the case of soft constraints, it is the inequality (or equality) itself which is viewed to be fuzzy (i.e. $Ax \leq b$ for a linear programming problem). This is distinct from the case in which the right hand side has vague value, in which case the right hand side is viewed to be fuzzy (i.e. $Ax \leq \tilde{b}$). To illuminate the difference between fuzzy inequalities and fuzzy right hand sides, consider the example of a computer dating service. Suppose woman A specifies that she would like to date a "medium-height" man. Woman A defines "medium" as a fuzzy set characterized by a triangular membership function, centered at 69", with a spread of 3". This is a hard constraint because the man is *required* to be medium-height, but medium-height is a fuzzy set. This is, therefore, an example of a fuzzy right-hand side (i.e. $Ax = \tilde{b}$). Now consider woman B, who desires to date a man of about 69" in height. Unlike woman A, woman B is willing to compromise a little on the height requirement in order to be matched with a date who meets some of her other requirements. Her satisfaction level with a 69" man is 1, with a 68" or 70" man is $\frac{2}{3}$, and with a 67" or 71" man is $\frac{1}{3}$. This is an example of a soft constraint, represented by a fuzzy equality (i.e. $Ax = b$). Notice that the membership functions in the two fuzzy cases are the same (symmetric triangular centered at 69" with a spread of 3"), but the semantics are different.

One result of the distinction between fuzziness and possibilistic uncertainty is that they manifest themselves in different regions of the optimization problem. Given the basic linear program,

$$\begin{aligned} \min z &= c^T x \\ \text{subject to: } Ax &\leq b, \\ x &\geq 0, \end{aligned} \quad (11)$$

fuzziness can occur in the right-hand-side ($\tilde{b}$), and/or in the inequality ($\leq$). To date, no model has been proposed which deals with fuzzy objective function coefficients of fuzzy constraint matrix coefficients. However, suppose a constraint, $a_{ij}x \leq b_i$, $i \in [1, n]$, is meant to apply only to members of a particular fuzzy set, $\tilde{Y}$. Now suppose that the element represented by row i of the constraint matrix is a member of set Y to a degree defined by $\mu_Y(y)$. Then the constraint i should be multiplied by the membership value of y_i , resulting in fuzzy coefficients. A similar argument can be applied for objective function coefficients. Possibilistic values can occur in the objective function coefficients $\hat{c}$, in the constraint matrix coefficients $\hat{A}$, and/or in the right-hand-side $\hat{b}$. To date, there is neither semantic nor model for a possibilistic inequality.

Fuzzy Relations

Optimization models usually involve constraints consisting of equalities, inequalities, or both. For deterministic optimization, the meaning and computation of the constraint set is clear. In optimization under uncertainty, however, the meaning and computation of “equality” and “inequality” must be determined. To this end, Dubois and Prade, (Dubois and Prade 1983; Dubois and Prade 1987), give a comprehensive analysis of fuzzy relations with four possible interpretations of fuzzy equalities, called modalities, and four possible interpretations of fuzzy inequalities. Inuiguchi (Inuiguchi et al. 1992a) adds two more modalities. However, only the original four modalities are outlined below, using fuzzy intervals $\tilde{M} = [m^-(\alpha), m^+(\alpha)]$, $0 \leq \alpha \leq 1$, and $\tilde{N} = [n^-(\alpha), n^+(\alpha)]$, $0 \leq \alpha \leq 1$.

The statement $\tilde{M} \geq \tilde{N}$ can be interpreted in any of the four following ways:

- (i) $\forall x \in \tilde{M}, \forall y \in \tilde{N}, x > y$. This is equivalent to $m^-(\alpha) > n^+(\alpha)$.

- (ii) $\forall x \in \tilde{M}, \exists y \in \tilde{N}, x \geq y$. This is equivalent to $m^-(\alpha) \geq n^-(\alpha)$.
 (iii) $\exists x \in \tilde{M}, \forall y \in \tilde{N}, x > y$. This is equivalent to $m^+(\alpha) > n^+(\alpha)$.
 (iv) $\exists (x, y) \in \tilde{M} \times \tilde{N}, x \geq y$. This is equivalent to $m^+(\alpha) \geq n^-(\alpha)$.

Inequality relation (i) indicates that x is necessarily greater than y , This is the pessimistic view. The decision maker who requires that $m^- > n^+$ in order to satisfy $\tilde{M} > \tilde{N}$ is taking no chances (Lodwick 1990a). (iv) indicates that x is possibly greater than y . This is the optimistic view. The decision maker who merely requires that $m^+ > n^-$ in order to satisfy $\tilde{M} > \tilde{N}$ has a hopeful outlook. Inequality relations (ii) and (iii) fall somewhere between the optimistic and pessimistic views.

The statement $\tilde{M} = \tilde{N}$ can be interpreted in any of the following four ways:

- (i) Zadeh's fuzzy set equality: $\mu_M = \mu_N$
 (ii) $\forall x \in \tilde{M}, \exists y \in \tilde{N}, x = y$ (which is equivalent to $\tilde{M} \subseteq \tilde{N}$).
 (iii) $\forall y \in \tilde{N}, \exists x \in \tilde{M}, x = y$ (which is equivalent to $\tilde{N} \subseteq \tilde{M}$).
 (iv) $\exists (x, y) \in \tilde{M} \times \tilde{N}, x = y$ (which is equivalent to $\tilde{N} \cap \tilde{M} \neq \emptyset$).

Equality relation (i) indicates that x is necessarily equal to y (the pessimistic view), (iv) indicates that x is possibly equal to y (the optimistic view), and (ii) and (iii) fall somewhere in between.

Basics of Fuzzy Sets and Possibility Theory in Optimization

The general fuzzy optimization model (1), (2), has two parts – the *objective* (normative criteria), $\text{opt } \tilde{z} = f(\tilde{c}, x)$ (1), and the *constraints*, $x \in \tilde{X}$ (2). This corresponds to what Kacprzyk and Orlovski state (p. 50 in Kacprzyk and Orlovski 1987b):

The analysis of real decision making situation is virtually based on two types of information:

- information on feasible alternative decisions (options, choices, alternatives, variants, . . .),
- information making possible the comparison of alternative decisions with each other in terms of “better”, “worse”, “indifferent”, etc.

The set of “feasible alternative decisions” is defined by the constraints while “the comparison of alternative decisions with each other” is accomplished by the objective. When the objective is a *utility function* that takes a given fuzzy set (or possibility distribution) and maps it to a subset of the real numbers and the constraints are transformed into equations and/or inequalities, the optimization model is a mathematical programming model. “Mathematical programming problems can be considered as decision making problems in which preferences between alternatives are described by means of objective function(s) on a set of alternatives given by constraints in such a way that more preferable alternatives have higher values” (Ramik and Vlach 2002a).

A key component of any mathematical programming problem is the input data. In our radiation therapy planning example, the input data includes the amount of radiation required to kill a cancerous cell, the maximum radiation a healthy cell can tolerate, and how much radiation a beam at some angle will deposit at a particular location in the body. When the input parameters of a mathematical programming model are described by uncertainty distributions (membership functions or possibility distributions), the mathematical programming problem becomes a fuzzy or possibilistic optimization problem. Recall that the decision-maker’s flexibility is modeled by fuzzy relations, which results in fuzzy optimization. In addition, we sometimes see fuzzy and possibilistic parameters combined in the same problem, or possibilistic parameters occurring in the statement of fuzzy goals. These situations result in a mixed fuzzy and possibilistic optimization problem. We note that for the models considered here, while parameters, relations, and even the value of the objective function might be fuzzy or possibilistic, all decisions are “crisp”. In

practice, the solution to a mathematical programming problem is useless if it is not implementable.

There have been many superb surveys of the area of fuzzy optimization. Among all of these, the following are noted: (Buckley 1988a; Delgado et al. 1994; Inuiguchi 1997, 2007a, b; Inuiguchi and Ramik 2000; Inuiguchi et al. 1990, 1992b; Kacprzyk and Orlovski 1987b; Lai and Hwang 1992; Liu 2002; Luhandjula 1989; Ramik and Vlach 2002a; Rommelfanger 2004; Rommelfanger and Slowinski 1998; Sahinidis 2004; Untiedt 2006; Verdegay 1982; Werners 1995), and (Zimmermann 1983).

It is clear that the fuzzy optimization model (1), (2) is ill-defined. The resolution of this ill-definition depends upon the function space in which the fuzzy optimization problem is solved. To date, two types of function spaces in which fuzzy optimization is “housed” have been used.

1. Fuzzy Banach Spaces (see Diamond and Kloeden 1994a, b; Jamison 1998; Saito and Ishii 1998)
2. Real Euclidean Space $\mathbb{R}^n$ (all other approaches)

Each of these two approaches has its own way of mapping the associated transitional and/or information deficiency uncertainty onto an ordered field. This order is necessary for the determination of optimality given the inherent normativeness of optimization.

Key Issues for Fuzzy Optimization Mapped to Fuzzy Banach Spaces

Methods for solving optimization problems often involve iteration and/or approximation. For fuzzy optimization problems, a Banach space facilitates the convergence analysis of iterative or approximation algorithms. When each successive iterate or approximation remains a fuzzy entity in a fuzzy Banach space, the convergent entity or approximation is also in the fuzzy Banach space – it is again a fuzzy entity. Once convergence is achieved, a decision can be based on this fuzzy entity. When the problem is solved in a fuzzy Banach space, the fuzzy sets in

the optimization problem remain fuzzy sets throughout the solution process and no translation is necessary except that of inclusion into an appropriate Banach space. Diamond and Kloeden (1994a, b) use a space in which their set of fuzzy sets, which they denoted E^n , over $\mathbb{R}^n$ is endowed with a neighborhood system and metric that renders it a Banach space. They then find the Karush–Kuhn–Tucker (KKT) conditions for optimality (see Diamond and Kloeden 1994b). Saito and Ishii (1998) also develop the KKT conditions for optimization over fuzzy numbers. Diamond (1991) and Jamison (1998, 2000) consider equivalence classes in developing a Banach space of fuzzy sets. Jamison (1998) goes on to use the Banach space developed from the equivalence classes to solve optimization problems. After the fuzzy optimization problem embedded in the fuzzy Banach space is solved, crisp decisions must be made based on fuzzy solution. This mapping from a fuzzy solution to the decision space is often called *defuzzification*. But up to the point of defuzzification all fuzzy entities from the model (1), (2) retain their fuzziness. Further discussion of fuzzy Banach space methods in fuzzy optimization is beyond the scope of this presentation.

Key Issues for Fuzzy Optimization Mapped to Real Euclidean Spaces The ill-definition of (1), (2) may be resolved by mapping the ambiguity and/or vagueness into a complete ordered lattice, namely, the real numbers. This is accomplished by translating the fuzzy objective and fuzzy constraint into real numbers or vectors, and real-valued relations. The various possible mappings distinguish among the types of fuzzy optimization. For a particular optimization problem in the form (1), (2), the answers to the following questions will determine which mapping is used.

1. Is the optimization fuzzy, possibilistic or a mixture?
2. What is meant by fuzzy/possibilistic function $f(\tilde{c}, x)$ and how does one compute $\tilde{z} = f(\tilde{c}, x)$?

3. What is meant by a fuzzy/possibilistic relation $x \tilde{\in} \tilde{X}$? How does one compute the resulting constraint set from fuzzy relations?
4. What is meant by fuzzy/possibilistic optimization of a fuzzy-valued function $\text{opt } \tilde{z}$?

Is the Optimization Fuzzy, Possibilistic, or a Mixture? The first issue is the nature of the uncertainty itself. This depends upon the semantics of the problem, as discussed in section “Semantics of Fuzzy Sets and Possibility Distributions in Fuzzy Optimization”. Recall that fuzzy entities are sets with non-sharp boundaries in which there is a transition between elements that belong and elements that don’t belong to the set. Possibilistic entities are known to exist but the evidence associated with whether a particular element belongs to the set or not is incomplete or hard to obtain.

With these definitions in mind, we briefly consider fuzzy, possibilistic, and mixed decision making. Much of what is presented next can be found in (Lodwick and Jamison 2005, 2007b).

Fuzzy Decision Making: Given the set of (crisp) decisions, Ω , and fuzzy sets, $\{\tilde{F}_i \mid i = 1 \text{ to } n\}$, find the optimal decision in the set Ω . That is,

$$\sup_{x \in \Omega} h(\tilde{F}_1(x), \dots, \tilde{F}_n(x)), \quad (12)$$

where $h: [0, 1]^n \rightarrow [0, 1]$ is an aggregation operator, often taken to be the min function, and $\tilde{F}_i(x) \in [0, 1]$ is the fuzzy membership of x in fuzzy set $\tilde{F}_i$. Note that the decision space Ω is a **crisp set** (a set of real numbers) and the optimal decision satisfies a mutual membership condition defined by the aggregation operator h . The methods of Bellman and Zadeh (1970), Tanaka, Okuda and Asai (1974), and Zimmermann (1974, 1976), who were the first to develop fuzzy mathematical programming fall into this category. While the aggregation operator h historically has been the min operator, it can be, for example, any t -norm that is consistent with the context of the problem and/or decision methods, for example, risk aversion (see Kaymak and Sousa 2003;

Rommelfanger 1994, or Werners 1988). For a discussion of aggregation operators see (Klir and Yuan 1995).

Possibilistic Decision Making: Given the set of (crisp) decisions, Ω , and the set of possibility distributions representing the uncertain outcomes from selecting decision $\vec{x} = (x_1, \dots, x_n)^T$ denoted $\Psi_x = \{\hat{F}_x^i, i = 1, \dots, n\}$, find the optimal decision that produces the best set of possible outcomes with respect to an ordering U of the outcomes. That is,

$$\sup_{\hat{F}_x^i \in \Psi} U(\hat{F}_x^1, \dots, \hat{F}_x^n), \quad (13)$$

where $U(\hat{F}_x^1, \dots, \hat{F}_x^n)$ represents a “utility” of the set of distributions of possible outcomes $\Psi = \{\Psi_x | x \in \Omega\}$. Note that the decision space Ψ is a set of (possibility) distributions $\hat{F}_x^i: \Omega \rightarrow [0, 1]$ resulting from taking decision $x \in \Omega$. This semantic is represented by the possibilistic optimization of (Inuiguchi 1992; Inuiguchi et al. 1992a, 1994; Jamison and Lodwick 2001).

Very simply, fuzzy decision making selects from a set of crisp elements ordered by an aggregation operator on corresponding membership functions while possibilistic decision making selects from a set of distributions measured by a utility operator that orders the corresponding distributions. These approaches have two different ordering operators (an aggregation operation like *min* for fuzzy sets and a utility function for possibility) which lead to different optimization methods (see Lodwick and Bachman 2005). The underlying sets associated with fuzzy decision making are fuzzy and the decision space consists of crisp elements from operations on these fuzzy sets. The underlying sets associated with possibilistic decision making are crisp and the decision space consists of distributions from operations on crisp sets.

Mixed Fuzzy/Possibilistic Decision Making: The issue of mixed fuzzy and possibility optimization problems has been studied as early as 1989 (see Delgado et al. 1989; Inuiguchi et al. 1990). In both these early models the same α -cut defines the level of ambiguity in the possibilistic coefficients

and the level at which the decision-maker’s requirements are satisfied. We interpret the solution to these models to mean that we have a possibility α of obtaining a solution that satisfies the decision maker to degree α . A recently proposed model (Untiedt 2007a) allows a trade-off between the fuzzy α -level and the possibilistic α -level. This allows the decision-maker to balance the likelihood of a solution with how satisfactory the solution is.

The use of distinct approaches for mathematical programming problems in which each constraint contains only one single kind of uncertainty is found in (Lodwick and Jamison 2005, 2007b). They aggregate all fuzzy constraint rows according to (12) and find a utility for the possibilistic rows according to (13). If a mixture of uncertainty occurs within one constraint, they apply interval-valued probability measure (IVPM) (Lodwick and Jamison 2006, 2008) to optimization. IVPM applied to optimization is discussed as a separate topic in section “Future Directions”.

What Is Meant by a Fuzzy-Valued Function and How Does One Compute $\tilde{z} = f(\tilde{c}, x)$? Any function of fuzzy sets (fuzzy-valued function) or possibility distributions must rely on an extension principle, as described in section “Extension Principles” (Lodwick 2007; Lodwick and Jamison 2003b), and independently (Baudrit et al. 2005). These authors present practical methods for computing the output of a fuzzy – or possibilistic-valued functions. For this presentation, it is assumed that (8) is used to evaluate fuzzy functions.

What Is Meant by a Fuzzy/Possibilistic Relation? There are several possible interpretations of fuzzy and possibilistic equalities and inequalities as detailed in section “Fuzzy Relations”. Other comparisons of fuzzy sets may be found in (Dubois and Prade 1983; Yager 1981). The choice of optimization model for a particular problem depends heavily on which of these relations is used.

What Is Meant by Fuzzy Optimization of a Fuzzy-Valued Function? There are three distinct kinds of optimality a fuzzy mathematical program might pursue. *First*, a program might seek to find,

among a collection of fuzzy sets, the optimal fuzzy set. This optimality depends on the order dictated by the fuzzy relations described in the previous section.

Secondly, a program may seek to maximize the membership function (or possibility) of the solution chosen for the problem. This typically happens when no fuzzy set satisfies the constraint at a full membership level. Note that $x \tilde{\in} \tilde{X}$ occurs in the transformation of the objective as well as the constraints. This is because some models allow constraint violations at a cost (to the objective function). There are several possible interpretations of fuzzy and possibilistic equalities and inequalities, as detailed in section “[Fuzzy Relations](#)”. The choice of optimization model for a particular problem depends heavily on which of these relations is used.

Thirdly, fuzzy optimization problems that use $\mathbb{R}^n$ as the basic space require a mapping of (1) and (2) from fuzzy sets and/or possibility distributions to real numbers and vectors. When the sense of optimization is fuzzy, $\widetilde{\text{opt}}$, optimization is interpreted as a goal, that is, the objective function is considered as a goal. Typically, a target is set and the objective is to come as close as possible to the target. For this presentation, $\widetilde{\text{opt}}$ will be considered as optimization over real numbers. Thus, the general fuzzy optimization problem is:

$$\widetilde{\text{opt}}_x \tilde{z} = f(\tilde{c}, x) \quad (14)$$

$$x \tilde{\in} \tilde{X}. \quad (15)$$

To transform (14) and (15) to $\mathbb{R}^n$, a mapping $T: \mathbb{R}^n \rightarrow \mathbb{R}^{n+1}$ is defined by:

$$\begin{aligned} \widetilde{\text{opt}}_x T \left(\begin{array}{c} f(\tilde{c}, x) \\ x \tilde{\in} \tilde{X} \end{array} \right) &= \widetilde{\text{opt}}_x \left[\begin{array}{c} T_1(f(\tilde{c}, x), x \tilde{\in} \tilde{X}) \\ T_2(x \tilde{\in} \tilde{X}) \end{array} \right] \\ &= \left[\begin{array}{c} \widetilde{\text{opt}}_x F(c, x) \in \mathbb{R} \\ x \in \Omega \subseteq \mathbb{R}^n \end{array} \right] \end{aligned} \quad (16)$$

Note that $x \tilde{\in} \tilde{X}$ occurs in the transformation of the objective as well as the constraints. This is because some fuzzy/possibilistic models consider violations of constraints as possible but at a penalty or cost (to the objective function). To make the discussion clearer, the general fuzzy optimization problem (1), (2) is restricted to the fuzzy linear programming model

$$\begin{aligned} \min \tilde{z} &= \tilde{c}^T x \\ \text{subject to: } \tilde{A}x &\leq \tilde{b} \\ x &\geq 0. \end{aligned} \quad (17)$$

In this context, the *objective* $\widetilde{\text{opt}} \tilde{z} = f(\tilde{c}, x)$ becomes

$$\min \tilde{z} = f(\tilde{c}, x) = \tilde{c}^T x. \quad (18)$$

The *constraint* $x \tilde{\in} \tilde{X}$ is

$$\tilde{X} = \{ \tilde{A}x \leq \tilde{b} \} \cup \{ x \geq 0 \}. \quad (19)$$

Interactivity To date, most fuzzy optimization models assume that the fuzzy and possibilistic parameters are non-interactive. Inuiguchi has studied fuzzy optimization with interactivity (dependencies) (Inuiguchi 2007a; Inuiguchi and Tanino 2004). The issue of interactivity is especially important for mathematical programming models in finance, such as portfolio models where groups of stocks are in fact known to be dependent. However, for this presentation, it is assumed that all uncertainties are non-interactive.

Classical Approaches to Fuzzy Optimization

There have been numerous formulations set forth for modeling imprecise objectives and constraints, each with associated input semantics and solution interpretations. This chapter organizes and reviews a representative selection of fuzzy and possibilistic formulations.

Fuzzy Programming

Vague parameter values leads to fuzzy programming. When the vagueness represents a willingness on the part of the decision-maker to bend the constraints, fuzzy programming can also be called flexible programming. In these cases, the decision-maker is willing to lower the feasibility requirements in order to obtain a more satisfactory objective function value, or, in some cases, simply in order to reach a feasible solution. Such flexible constraints are commonly referred to as *soft constraints*.

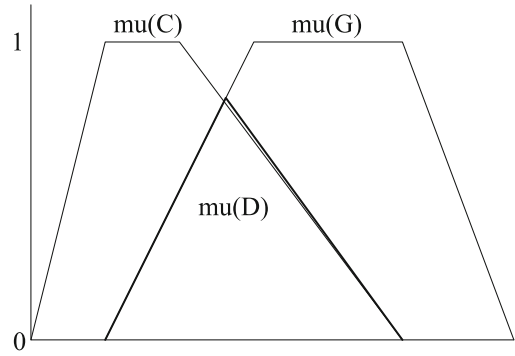
Bellman and Zadeh The landmark paper (Bellman and Zadeh 1970) by Bellman and Zadeh is the first to propose an approach to mathematical programming with fuzzy components. Conventional mathematical programming has three principal components: A set of alternatives, X , a set of constraints, C , which limit the choice of alternatives, and an objective function, z , which measures the desirability of each alternative. Bellman and Zadeh propose that, in a fuzzy environment, a more natural framework is one in which goal(s) replace the objective function(s), and a symmetry between goals and constraints erases the differences between them. (For this reason, flexible programming is sometimes called symmetric programming.)

Each fuzzy goal, $\tilde{G}_i$, or constraint, $\tilde{C}_j$, is defined as a fuzzy subset of the solution alternatives, X , via a membership function ($\mu_{G_i}(x)$ or $\mu_{C_j}(x)$). Then a fuzzy decision, $\tilde{D}$, is defined as a fuzzy set resulting from the intersection of $\tilde{C}$ and $\tilde{G}$, and is characterized by membership function μ_D as follows:

$$\begin{aligned}\mu_D(x) &= \mu_C(x) \wedge \mu_G(x) \\ &= \min[\mu_C(x), \mu_G(x)].\end{aligned}\quad (20)$$

An optimal decision, in turn, is one which maximizes μ_D . Figure 1 illustrates and provides a visual interpretation of μ_G , μ_C , and μ_D .

Tanaka, Okuda, and Asai (1974) Tanaka, Okuda, and Asai suggested an implementation



Fuzzy Optimization, Fig. 1 Fuzzy decision membership function μ_D

of Bellman and Zadeh's fuzzy decision using α -cuts (Tanaka et al. 1974). In the literature, α -cut, α -level, and α -set are used synonymously. If $\tilde{C}$ is a fuzzy constraint in X , then an α -cut of $\tilde{C}$, denoted by C_α , is the following crisp set in X :

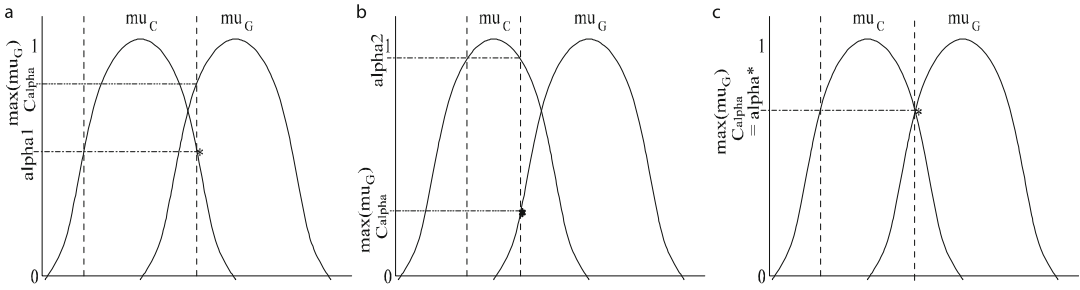
$$\begin{aligned}C_\alpha &= \{x | \mu_C(x) \geq \alpha\} \quad \text{for } \alpha \in (0, 1] \\ C_\alpha &= \text{cls}\{x | \mu_C(x) > 0\} \quad \text{for } \alpha = 0\end{aligned}\quad (21)$$

where $\text{cls}\{A\}$ denotes the closure of set $\{A\}$. Tanaka, Okuda and Asai (1974) make the following remarkable observation:

$$\sup_x \mu_D(x) = \sup_\alpha \left[\alpha \wedge \max_{C_\alpha} \mu_G(x) \right].$$

For illustration, consider two fuzzy sets, $\tilde{C}$ and $\tilde{G}$, depicted in Fig. 2.

In case (i), $\alpha = \alpha_1$, and C_α is the interval between the end-points of the α -cut, $[C^-(\alpha_1), C^+(\alpha_1)]$. The maximum μ_G in this interval is shown in the example. In this case, $\alpha_1 < \max_{C_{\alpha_1}} \mu_G(x)$, so $[\alpha_1 \wedge \max_{C_{\alpha_1}} \mu_G(x)] = \alpha_1$. In case (ii), $\alpha_2 > \max_{C_{\alpha_2}} \mu_G(x)$, so $[\alpha_2 \wedge \max_{C_{\alpha_2}} \mu_G(x)] = \max_{C_{\alpha_2}} \mu_G(x)$. In case (iii), $\alpha^* = \max_{C_{\alpha^*}} \mu_G(x)$. It should be apparent from Fig. 2 that $\alpha^* = \sup \mu_D$. In case (iii), $\alpha = \alpha^*$ is also $\sup_\alpha [\alpha \wedge \max_{C_\alpha} \mu_G(x)]$. For any $\alpha < \alpha^*$, we have case (i), where $[\alpha \wedge \max_{C_\alpha} \mu_G(x)] = \alpha < \alpha^*$; and for any $\alpha > \alpha^*$, we have case (ii), where $[\alpha \wedge \max_{C_\alpha} \mu_G(x)] = \max_{C_\alpha} \mu_G(x) < \alpha^*$. The



Fuzzy Optimization, Fig. 2 Illustration of a Bellman/Zadeh fuzzy decision set

formal proof, which follows the reasoning illustrated here pictorially, is omitted for brevity's sake. However, the interested reader is referred to Tanaka, Okuda, and Asai (1974).

This result allows the following reformulation of the fuzzy mathematical problem:

Determine (α^*, x^*)

$$\alpha^* \wedge f(x^*) = \sup_{\alpha} \left[\alpha \wedge \max_{x \in C_{\alpha}} f(x) \right]. \quad (22)$$

The researchers suggest an iterative algorithm which solves the problem. The algorithm cycles through a series of steps, each of which brings it closer to a solution to the relation $\alpha^* = \max_{C_{\alpha}} f(x)$. When α^* is determined to within a tolerable degree of uncertainty, it is used to find x^* such that $f(x^*) = \max_{C_{\alpha^*}} f(x)$.

This cumbersome solution method is impractical for large-scale problems. It should be noted that Negoita (Negoita and Sularia 1976) suggested an alternative, but similarly complex, algorithm for the flexible programming problem based on Tanaka's findings.

Zimmermann Just two years after Tanaka, Okuda, and Asai (1974) suggested the use of α -cuts to solve the fuzzy mathematical problem, Zimmermann published a linear programming equivalent to the α -cut formulation.

Beginning with the crisp programming problem,

$$\begin{aligned} \min Z &= cx \\ \text{subject to: } Ax &\leq b \\ x &\geq 0, \end{aligned} \quad (23)$$

the decision maker introduces flexibility in the constraints, and sets a target value for the objective function, Z ,

$$\begin{aligned} cx &\leq Z \\ Ax &\leq b \\ x &\geq 0 \end{aligned} \quad (24)$$

A linear membership function μ_i is defined for each flexible right hand side, including the goal Z as

$$\mu_i \left(\sum_j a_{ij} x_j \right) = \begin{cases} 1 & \sum_j a_{ij} x_j \leq b_i, \\ \frac{1 - \left(\sum_j a_{ij} x_j - b_i \right)}{d_i} & \sum_j a_{ij} x_j \in (b_i, b_i + d_i), \\ 0 & \sum_j a_{ij} x_j \geq b_i + d_i. \end{cases}$$

where d_i is the decision maker's maximum allowable violation of constraint (or goal) i .

According to the convention set forth by Bellman and Zadeh, the fuzzy decision D is characterized by

$$\mu_D = \min_i \mu_i \left(\sum_j a_{ij} x_j \right), \quad (25)$$

and

$$\max_{x \geq 0} \min_i \mu_i \left(\sum_j a_{ij} x_j \right) \quad (26)$$

is the decision with the highest degree of membership.

The problem of finding the solution is therefore

$$\begin{aligned} & \max \mu_D(x) \\ & \text{subject to: } x \geq 0. \end{aligned} \quad (27)$$

In the membership functions μ_i from (25), Zimmermann substitutes b'_i for b_i/d_i and B'_i for $\sum_j a_{ij}/d_i$. He also drops the 1 (which does not change the solution to the problem) to obtain the simplification $\mu_i = b'_i - B'_i x$. Equation (27) then becomes

$$\begin{aligned} & \min_i (b'_i - B'_i x) \\ & x \geq 0. \end{aligned} \quad (28)$$

This is equivalent to the following linear program:

$$\begin{aligned} & \max \alpha \\ & \alpha \leq b'_i - B'_i x \quad \forall i \\ & x \geq 0, \quad 0 \leq \alpha \leq 1. \end{aligned} \quad (29)$$

Equation (29) can easily be solved via the simplex or interior point methods.

Verdegay The solutions examined so far for the fuzzy programming problem have been crisp solutions. Ralescu (1977) first suggested that a fuzzy problem should have a fuzzy solution and Verdegay (1982) proposes a method for obtaining a fuzzy solution. Verdegay considers a problem with fuzzy constraints,

$$\begin{aligned} & \max z = f(x) \\ & \text{subject to: } x \subseteq \tilde{C}, \end{aligned} \quad (30)$$

where the set of constraints have a membership function $\mu_{\tilde{C}}$, with alpha-cuts $\tilde{C}_\alpha$.

Verdegay defines x_α as the set of solutions that satisfy constraints $\tilde{C}_\alpha$. Then a fuzzy solution to the fuzzy linear programming problem is

$$\begin{aligned} & \max_{x \in \tilde{C}_\alpha} z = f(x) \\ & \forall \alpha \in [0, 1]. \end{aligned} \quad (31)$$

Verdegay proposes solving (31) parametrically for $\alpha \in [0, 1]$ to obtain a fuzzy solution $\tilde{X}$, with α - cut x_α , which yields fuzzy objective value $\tilde{z}$, with α - cut z_α .

It should be noted that the (crisp) solution obtained via Zimmermann's method corresponds to Verdegay's solution in the following way: If the objective function is transformed into a goal, with membership function μ_G , then Zimmermann's optimal solution, x^* , is equal to Verdegay's optimal value $x(\alpha)$ for the value of α which satisfies

$$\mu_G(z_\alpha^* = cx_\alpha^*) = \alpha.$$

In other words, when a particular α - cut of the fuzzy solution, (x_α) yields an objective value $(z_\alpha = c^T x_\alpha)$ whose membership level for goal G , $(\mu_G(z_\alpha))$ is equal to the same α , then that solution x_α corresponds to Zimmermann's optimal solution, x^* .

Surprise Recently, Neumaier suggested a way to model fuzzy right-hand side values with crisp inequality constraints $(Ax \leq \tilde{b})$ based on the concept of surprise (Lodwick et al. 2001; Neumaier 2003). Neumaier defines a surprise function, $s(x | E)$, which corresponds to the amount of surprise a variable x produces, given a statement E . The range of s is $[0, \infty)$, with $s = 0$ corresponding to an entirely true statement E , and $s = \infty$ corresponding to an entirely false statement E .

The most plausible values are those values of x for which $s(x | E)$ is minimal, so Neumaier proposes that the best compromise solution for an optimization problem with fuzzy goals and constraints can be found by minimizing the sum of the surprise functions of the goals and constraints. Each fuzzy constraint,

$$(Ax)_i \leq \tilde{b}_i,$$

is translated into a fuzzy equality constraint,

$$(Ax)_i = \tilde{\xi}_i,$$

where the membership function $\mu_i(\xi)$ of $\tilde{\xi}_i$ is the possibility that $\tilde{b}_i \geq \xi$. These membership functions are subsequently translated into surprise functions by

$$s_i(\xi) = \left(\mu_i(\xi)^{-1} - 1 \right)^2;$$

and the contribution of all constraints are added to give the total surprise

$$\sum_i s_i(\xi) = \sum_i s_i((Ax)_i).$$

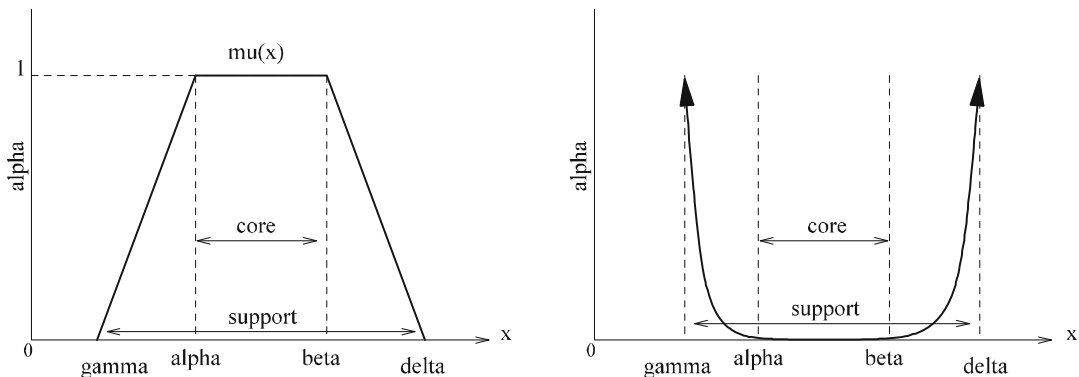
Thus, the fuzzy optimization problem is

$$\begin{aligned} \min z &= \sum_i s_i((Ax)_i) \\ \text{subject to: } x &\geq 0. \end{aligned} \quad (32)$$

Figure 3 illustrates a surprise function associated with a corresponding fuzzy goal. For triangular and trapezoidal numbers, the surprise function is simple, smooth, and convex, leading to a tractable non-linear programming problem.

The surprise approach is a fuzzy optimization method since the optimization is over sets of crisp values derived from fuzzy sets. In contrast to flexible programming, the constraints are not restricted such that **all** satisfy a minimal level. The salient feature is that surprise uses a dynamic penalty for falling outside distribution/member-ship values of one. Because the individual penalties are convex functions which become infinite as the values approach the endpoints of the support, this method lends itself to convex programming solution techniques.

It is noted that the surprise approach may be used to handle soft constraints of flexible programming since these soft constraints can be viewed as fuzzy numbers (trapezoidal fuzzy numbers when linear interpolation is used). However, if soft constraints are handled using surprise functions, the **sum** of the failure to meet the constraints is minimized rather than forcing each constraint to meet a minimal feasibility level. One could add a hard constraint to the surprise approach to attain this minimal level of feasibility. The modeler might choose to translate soft constraints to surprise functions (with perhaps a fixed minimal feasibility level) because surprise is usually more computationally efficient than Zimmermann's method of handling soft constraints (see Lodwick and Bachman 2005). Therefore, the surprise



Fuzzy Optimization, Fig. 3 Surprise function

approach is quite flexible both in terms of semantics as well as in computational robustness.

Possibilistic Programming

Recall that fuzzy imprecision arises when elements of a set (for instance, a feasible set) are members of the set to varying degrees, which are defined by the membership function of the set. Possibilistic imprecision arises when elements of a set (say, again, a feasible set) are known to exist as either full members or non-members, but whether they are members or non-members is known with a varying degree of certainty, which is defined by the possibility distribution of the set. This possibilistic uncertainty arises from a lack of information. In this section, we examine possibilistic programming formulations.

Buckley J.J. Buckley has suggested an algorithm for dealing with possibilistic cost coefficients, constraint coefficients, *and* right hand sides. Consider the possibilistic linear program

$$\begin{aligned} \min Z &= \hat{C}x \\ \text{subject to: } \hat{A}x &\geq \hat{b}, x \geq 0. \end{aligned} \quad (33)$$

where $\hat{A} = [\hat{a}_{ij}]$ is an $m \times n$ matrix of trapezoidal possibilistic intervals $\hat{a}_{ij} = (a_{ij\alpha}, a_{ij\beta}, a_{ij\gamma}, a_{ij\delta})$, $\hat{b} = (\hat{b}_1, \dots, \hat{b}_m)^T$ is an $m \times 1$ vector of trapezoidal fuzzy numbers $\hat{b}_i = (b_{i\alpha}, b_{i\beta}, b_{i\gamma}, b_{i\delta})$, and $\hat{C} = (\hat{c}_1, \dots, \hat{c}_n)$ is a $1 \times n$ vector of trapezoidal possibilistic intervals $\hat{c}_i = (c_{i\alpha}, c_{i\beta}, c_{i\gamma}, c_{i\delta})$. The possibilistic intervals are the possibility distributions associated with the variables, and place a restriction on the possible values the variables may assume (Zadeh 1978). For example, $\text{Poss}[\hat{a}_{ij} = a] = \mu_a(\tilde{a}_{ij})$ is the possibility that $\tilde{a}_{ij}$ is equal to a . Stated another way, $\text{Poss}[\hat{a}_{ij} = a] = x$ means that, given the current state of knowledge about the value of a_{ij} , we believe that x is the possibility that variable a_{ij} could take on value a .

Because this is a possibilistic linear programming problem, the objective function will be governed by a possibilistic distribution, $\text{Poss}[Z = z]$.

Let us consider this simple example as we follow Buckley's derivation of $\text{Poss}[Z = z]$:

$$\begin{aligned} \min z &= \tilde{d}x_1 + \tilde{e}x_2 \\ \text{subject to } \tilde{f}x_1 + \tilde{g}x_2 &\leq \tilde{h} \\ \tilde{i}x_1 + \tilde{j}x_2 &\leq 0 \\ x_1, x_2 &\geq 0 \end{aligned} \quad (34)$$

To derive the possibility function, $\text{Poss}[Z = z]$, Buckley first specifies the possibility that x satisfies the i th constraint. Let

$$\Pi(\hat{a}_i, \hat{b}_i) = \min(\mu_{a_{i1}}(\tilde{a}_{i1}), \dots, \mu_{a_{im}}(\tilde{a}_{im}), \mu_{b_i}(\tilde{b}_i)), \quad (35)$$

which is the simultaneous distribution of $\tilde{a}_{ij}$, $1 \leq j \leq n$, and $\tilde{b}_i$. In our example (34) this corresponds to

$$\Pi(\hat{f}, \hat{g}, \hat{h}) = \min(\mu_F(f), \mu_G(g), \mu_H(h)).$$

Then the possibility that $x \geq 0$ is feasible with respect to the i th constraint is:

$$\text{Poss}[x \in \mathcal{F}_i] = \sup_{a_i, b_i} (\Pi(a_i, b_i) | a_i x \geq b_i).$$

In our example (34) this corresponds to

$$\begin{aligned} \text{Poss}[x \in \mathcal{F}_1] &= \sup_{f, g, h} (\Pi(f, g, h) | \tilde{f}x_1 + \tilde{g}x_2 \leq h) \\ \text{Poss}[x \in \mathcal{F}_2] &= \sup_{i, j, k} (\Pi(i, j, k) | \tilde{i}x_1 + \tilde{j}x_2 \leq k). \end{aligned} \quad (36)$$

Now the possibility that $x \geq 0$ is feasible with respect to *all* constraints is

$$[x \in \mathcal{F}] = \min_{1 \leq i \leq m} ([x \in \mathcal{F}_i])$$

In our example (34),

$$[x \in \mathcal{F}] = \min([x \in \mathcal{F}_1], [x \in \mathcal{F}_2])$$

Buckley next constructs $\text{Poss}[Z = z | x]$, which is the conditional possibility that the objective

function Z obtains a particular value z , given values for x . The joint distribution of the possibilistic cost coefficients $\tilde{c}_j$ is

$$\Pi(c) = \min(\mu_{c_1}(\tilde{c}_1), \dots, \mu_{c_n}(\tilde{c}_n)). \quad (37)$$

In our example (34),

$$\Pi(c) = \min(\mu_D(\tilde{d}), \mu_E(\tilde{e})).$$

Therefore

$$\text{Poss}[Z = z | x] = \sup_c (\Pi(c) | cx = z). \quad (38)$$

In our example (34), to obtain the possibility that the objective function will take on a particular value z , given the solution x , (i.e. $\text{Poss}[Z = z | x]$), we consider all the values of d and e for which $dx_1 + ex_2 = z$. Of these, the pair with the highest simultaneous possibility values will yield an objective function value of z with the same possibility level. For example, consider the case illustrated in Fig. 4.

When $\tilde{d} = (2, 2)$ and $\tilde{e} = (3, 2)$ are symmetric triangular fuzzy numbers, $(x_1, x_2) = (1, 1)$ and $z = 7$. We consider values of d and e such that $d \times 1 + e \times 1 = 7$. We could select $d = 2$ and $e = 5$ with $\mu_D(2) = 1$ and $\mu_E(5) = 0$. The joint possibility of $d = 2$ and $e = 5$ is $\min(\mu_D(2), \mu_E(5)) = 0$. The maximum joint possibility of d and e such that $d \times 1 + e \times 1 = 7$ occurs when $d = 3$ and $e = 4$. The joint possibility that $d = 3$ and $e = 4$, which

is $\min(\mu_D(3), \mu_E(4)) = .5$, we set equal to the possibility that $z = 7$ given that $(x_1, x_2) = (1, 1)$. So $\text{Poss}[z = 7 | (1, 1)] = .5$. A combination of (37) and (38) yields the possibility distribution of the objective function:

$$\text{Poss}[Z = z] = \sup_{x \geq 0} [\min(\text{Poss}[Z = z | x], \text{Poss}[x \in \mathcal{F}])]. \quad (39)$$

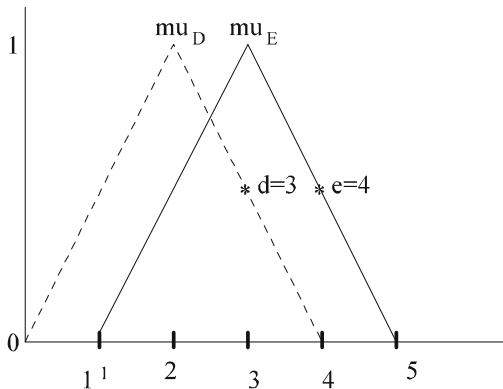
In our example (34), to obtain the possibility that the objective function will take on a particular value z (that is $\text{Poss}[Z = z]$, we consider $\text{Poss}[Z = z | x]$, as described above, for all positive x 's, and select the x which maximizes $\text{Poss}[Z = z | x]$.

This definition of the possibility distribution motivates Buckley's solution method. Recall that because we are dealing with a possibilistic problem, the solution will be governed by a possibilistic distribution. Buckley's method depends upon a static α , chosen a priori. The decision maker defines an acceptable level of uncertainty in the objective outcome, $0 < \alpha \leq 1$. For a given α , we define the left and right end-points of the α -cut of a fuzzy number $\tilde{x}$ as $x^-(\alpha)$ and $x^+(\alpha)$, respectively. Using these, Buckley defines a new objective function:

$$\begin{aligned} \min Z(\alpha) &= c^-(\alpha)x \\ \text{subject to: } &A^+(\alpha)x \geq b^-(\alpha). \end{aligned} \quad (40)$$

Since both x and α are variables, this is a non-linear programming problem. When α is fixed in advance, however, it becomes linear. We can use either the simplex method or an interior point method to solve for a given a priori chosen value of α . If a maximal value for α is desired, the linear programming method must be applied iteratively.

It should be noted that this linear program is constrained upon the best-case scenario. That is, for a given α - level, each variable is multiplied by the largest possible coefficient $a_{ij}^+(\alpha)$, and is required to be greater than the the smallest possible right hand side $b_i^-(\alpha)$. We should interpret $z(\alpha)$ accordingly. If the solution to the linear program is implemented, the possibility that the objective function will attain the level $z(\alpha)$ is given by α . Stated another way, the best-case scenario is that



Fuzzy Optimization, Fig. 4 Buckley example

the objective function attains a value of $z(\alpha)$, and the possibility of the best case scenario occurring is α .

Tanaka, Asai, Ichihashi In the mid 1980s, Tanaka and Asai (1984) and Tanaka, Ichahashi, and Asai (1984) proposed a technique for dealing with ambiguous coefficients and right hand sides based upon a possibilistic definition of “greater than zero”. The reader will note that this approach bears many similarities to the flexible programming proposed by Tanaka, Okuda, and Asai a decade earlier, which was discussed in subsection “Tanaka, Okuda, and Asai”. Indeed, the 1984 publications refer to *fuzzy* variables. This approach has subsequently been classified (Inuiguchi et al. 1990; Lai and Hwang 1992) as possibilistic programming because the imprecision it represents stems from a lack of information about the values of the coefficients.

Consider a programming problem with non-interactive possibilistic $\hat{A}$, $\hat{b}$, and $\hat{c}$, whose possible values are defined by fuzzy matrix $\tilde{A}$, fuzzy vector $\tilde{b}$, and fuzzy vector $\tilde{c}$, respectively:

$$\begin{aligned} \min \tilde{z} &= \tilde{c}x \\ \text{subject to: } \tilde{A}x &\leq \tilde{b} \\ x &\geq 0. \end{aligned} \quad (41)$$

Tanaka, Asai, and Ichihashi transform the problem in several steps. First, the objective function is viewed as a goal. As in flexible programming, the goal becomes a constraint, with the aspiration level for the objective function on the right-hand-side of the inequality. Next, a new decision variable x_0 is added. Finally, the b ’s are moved to the left hand side, so that the possibilistic linear programming problem is

$$\begin{aligned} \tilde{a}'_i x'_i &\geq 0, \quad i = 1, 2, \dots, m \\ x' &\geq 0 \end{aligned} \quad (42)$$

where $x' = (1, x^T)^T = (1, x_1, x_2, \dots, x_n)^T$, and $\tilde{a}'_i = (\tilde{b}_i, \tilde{a}_{i1}, \dots, \tilde{a}_{in})$.

Note that all the parameters, A , b , and c are now grouped together in the new constraint matrix $\tilde{A}$. Because the objective function(s) have become goals, the cost coefficients, c , are part of the constraint coefficient matrix A . Furthermore, because the right-hand side values have been moved to the left-hand side, the b ’s are also part of the constraint coefficient matrix. The right-hand-sides corresponding the former objective functions are the aspiration levels of the goals.

Each constraint becomes

$$\tilde{Y}_i = \tilde{a}_i x \geq 0,$$

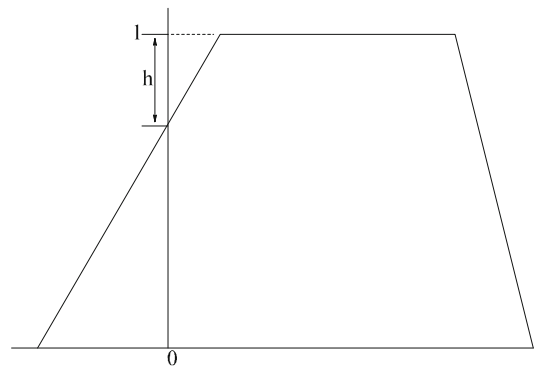
where

$$\tilde{a}_i = (\tilde{b}_i, \tilde{a}_{i1}, \dots, \tilde{a}_{in}).$$

“ $\tilde{Y}_i$ is almost positive”, denoted by $\tilde{Y}_i \gtrsim 0$, is defined by

$$\begin{aligned} \tilde{Y}_i &\gtrsim 0 \\ &\Leftrightarrow \\ \mu_{Y_i}(0) &\leq 1 - h, \\ x^T \alpha_i &\geq 0. \end{aligned} \quad (43)$$

The measure of the non-negativity of $\tilde{Y}_i$ is h : The greater the value of h , the stronger the meaning of “almost positive” (see Fig. 5). Actually, h is $1 - \alpha$, where α is the level of membership used by Bellman and Zadeh.



Fuzzy Optimization, Fig. 5 h-level

Tanaka and Asai (1984) developed this theory for triangular fuzzy numbers, and Tanaka and Asai extended it to trapezoidal fuzzy numbers in (Tanaka et al. 1985). Inuiguchi, Ichihashi, and Tanaka, (1990) generalized the approach for L - R fuzzy numbers. For the sake of simplicity, our discussion here will deal with trapezoidal fuzzy numbers, denoted $\tilde{x} = (\gamma, \alpha, \beta, \delta)$, where (γ, δ) is the support of $\tilde{x}$, and (α, β) is the core of $\tilde{x}$.

Using (43), we can rewrite each constraint from (42) as

$$\begin{aligned} \mu_{Y_i}(0) &= 1 - \frac{\alpha_i^T x}{(\alpha_i - \gamma_i)^T x} \\ &\leq 1 - h \\ \alpha_i^T x &\geq 0, \end{aligned} \quad (44)$$

where $x > 0$. Then (44) reduces to

$$(\alpha_i - h(\alpha_i - \gamma_i))^T x \geq 0.$$

Since we wish to find the largest h that satisfies these conditions, the linear program becomes

$$\begin{aligned} \max z &= h \\ \text{subject to: } &(\alpha_i - h(\alpha_i - \gamma_i))^T x \geq 0, \quad \forall i \quad (45) \\ &h \in [0, 1]. \end{aligned}$$

Since both x and h are variables, this is a non-linear programming problem. When h is fixed, it becomes linear. We can use the simplex method or an interior point algorithm to solve for a given

value of h . If the decision maker wishes to maximize h , the linear programming method must be applied iteratively.

Fuzzy Max Dubois and Prade (1980a) suggested that the concept of “fuzzy max” could be applied to constraints with fuzzy parameters. The “fuzzy max” concept was used to solve possibilistic linear programs with triangular possibilistic coefficients by Tanaka, Ichihashi, and Asai (1984). Ramik and Rimanek (1985) applied the same technique to L - R fuzzy numbers. For consistency, we will discuss the fuzzy max technique with respect to trapezoidal numbers.

The fuzzy max, illustrated in Fig. 6 where $C = \max[A, B]$, is the extended maximum operator between real numbers, and defined by the extension principle (8) as

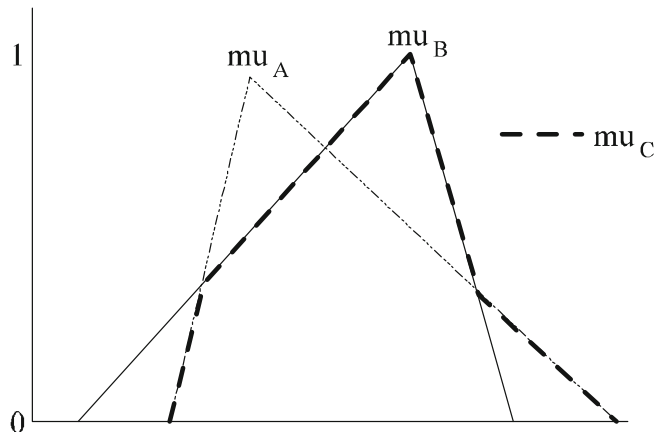
$$\mu_C(c) = \max_{\{a, b: c = \max(a, b)\}} \min[\mu_A(a), \mu_B(b)].$$

Using fuzzy max, we can define an inequality relation as

$$\begin{aligned} \tilde{A} &\geq \tilde{B} \\ \Leftrightarrow \\ \max(\tilde{A}, \tilde{B}) &= \tilde{A}. \end{aligned} \quad (46)$$

Applying (46) to a fuzzy inequality constraint:

Fuzzy Optimization,
Fig. 6 Illustration of fuzzy
max



$$\begin{aligned}
 f(x, \tilde{a}) &\geq g(x, \tilde{b}) \\
 &\Leftrightarrow \\
 \max(f(x, \tilde{a}), g(x, \tilde{b})) &= f(x, \tilde{a}).
 \end{aligned} \tag{47}$$

Observe that the inequality relation defined by (46) yields only a partial ordering. That is, it is sometimes the case that neither $\tilde{A} \geq \tilde{B}$ nor $\tilde{B} \geq \tilde{A}$ holds. To improve this, Tanaka, Ichihashi, and Asai, introduce a level h , corresponding to the decision maker's degree of optimism. They define an h -level set of $\tilde{A} \geq \tilde{B}$ as

$$\begin{aligned}
 \tilde{A} \geq_h \tilde{B} \\
 &\Leftrightarrow \\
 \max [\tilde{A}]_\alpha &\geq \max [\tilde{B}]_\alpha \\
 \min [\tilde{A}]_\alpha &\geq \min [\tilde{B}]_\alpha \\
 \alpha &\in [1 - h, 1].
 \end{aligned} \tag{48}$$

This definition of $\tilde{A} \geq \tilde{B}$ requires that two of Dubois' inequalities from section "Fuzzy Relations", (ii) and (iii), hold at the same time, and is illustrated in Fig. 7.

Tanaka, Ichihashi, and Asai, (1984) suggest a similar treatment for a fuzzy objective function. A problem with the single objective function, Maximize $z(x, \tilde{c})$, becomes a multi-objective problem with objective functions:

$$\text{maximize} \begin{cases} \inf(z(x, \tilde{c})_\alpha) & \alpha \in [0, 1] \\ \sup(z(x, \tilde{c})_\alpha) & \alpha \in [0, 1]. \end{cases} \tag{49}$$

Clearly, since α can assume an infinite number of values, (49) has an infinite number of parameters. Since (49) is not tractable, Inuiguchi, Ichihashi, and Tanaka (1990) suggest using the following approximation using a finite set of $\alpha_i \in [0, 1]$:

$$\text{maximize} \begin{cases} \min(z(x, \tilde{c}))_{\alpha_i} & i = 1, 2, \dots, p \\ \max(z(x, \tilde{c}))_{\alpha_i} & i = 1, 2, \dots, p \end{cases} \tag{50}$$

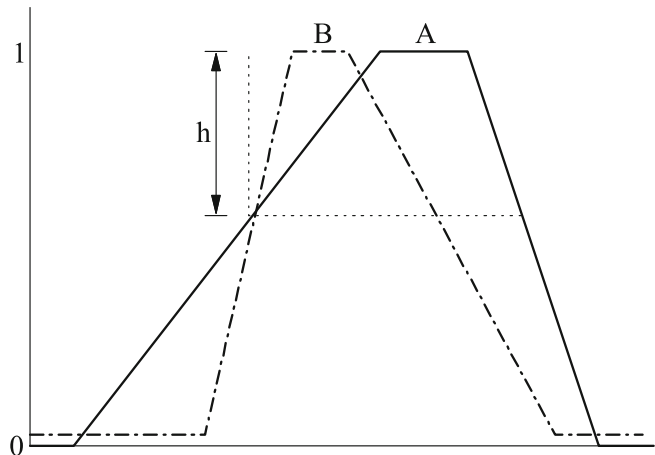
Jamison and Lodwick Jamison and Lodwick (2001; Lodwick and Jamison 1997) develop a method for dealing with possibilistic right hand sides that is a possibilistic generalization of the recourse models in stochastic optimization. Violations of constraints are allowable, at a cost determined a priori by the decision maker.

Jamison and Lodwick choose the utility (that is, valuation) of a given interval of possible values to be its expected average (a concept defined by Yager (1980).) The expected average (or EA) of a possibilistic distribution of $\tilde{a}$ is defined to be

$$EA(\tilde{a}) = \frac{1}{2} \int_0^1 (\tilde{a}^-(\alpha) + \tilde{a}^+(\alpha)) d\alpha. \tag{51}$$

Fuzzy Optimization,

Fig. 7 Illustration of $\tilde{A} \geq \tilde{B}$



It should be noted that the expected average of a crisp value is the value itself, since $\tilde{a}^-(\alpha) = \tilde{a}^+(\alpha) = a$, $EA(a) = \frac{1}{2} \int_0^1 (a + a) d\alpha = a \int_0^1 d\alpha = a$.

Jamison and Lodwick start from the following possibilistic linear program:

$$\begin{aligned} \max z &= c^T x \\ Ax &\leq \hat{b}, \\ x &\geq 0. \end{aligned} \quad (52)$$

By subtracting a penalty term from the objective function, they transform (52) into the following possibilistic non-linear program:

$$\begin{aligned} \max z &= c^T x + p^T \max(0, Ax - \hat{b}) \\ x &\geq 0. \end{aligned} \quad (53)$$

The “max” in the penalty is taken component-wise and each $p_i < 0$ is the cost per unit violation of the right hand side of constraint i . The utility, which is the expected average, of the objective function is chosen to be minimized. The possibilistic programming problem becomes

$$\begin{aligned} \max z &= c^T x + pEA\left(\max(0, Ax - \hat{b})\right) \\ x &\in [0, U]. \end{aligned} \quad (54)$$

A closed-form objective function (for the purpose of differentiating when solving) is achieved in (Lodwick and Bachman 2005) by replacing

$$\max(0, Ax - \hat{b})$$

with

$$\frac{\sqrt{(Ax - \hat{b})^2 + \epsilon^2} + Ax - \hat{b}}{2}.$$

Jamison and Lodwick’s method can be extended, (Jamison and Lodwick 2001), to account for possibilistic values for A , b , c , and even the penalty coefficients p with the following formulation:

$$\begin{aligned} EA\tilde{f}(x) &= \frac{1}{2} \int_0^1 \left\{ \tilde{c}^-(\alpha)^T x + \tilde{c}^+(\alpha)^T x \right. \\ &\quad - \left[\tilde{p}^+(\alpha) \max(0, \tilde{A}^-(\alpha)x - \tilde{b}^+(\alpha)) \right] \\ &\quad \left. - \left[\tilde{p}^-(\alpha) \max(0, \tilde{A}^+(\alpha)x - \tilde{b}^-(\alpha)) \right] \right\} d\alpha. \end{aligned} \quad (55)$$

This approach differs significantly from the others we’ve examined in several regards. First, many of the approaches we’ve seen have incorporated the objective function(s) as goals into the constraints in the Bellman and Zadeh tradition. Jamison and Lodwick, on the other hand, incorporate the constraints into the objective function. Bellman and Zadeh create a symmetry between constraints and objective, while Jamison and Lodwick temper the objective with the constraints. A second distinction of the expected average approach is the nature of the solution. The other formulations we have examined to this point have produced either (1) a crisp solution for a particular value of α , (namely, the maximal value of α), or, (2) a fuzzy/possibilistic solution which encompasses all possible α values. The Jamison and Lodwick approach provides a crisp solution via the expected average utility which encompasses all alpha values. This may be a desirable quality to the decision maker who wants to account for all possibility levels and still reach a crisp solution.

Luhandjula Luhanjula’s (1986) formulation of the possibilistic mathematical program depends upon his concept of “more possible” values. He first defines a possibility distribution Π_X with respect to constraint F as

$$\Pi_X = \mu_F(u),$$

where $\mu_F(u)$ is the degree to which the constraint F is satisfied when u is the value assigned to the solution X .

Then the set of more possible values for X , denoted by $V_p(X)$, is given by

$$V_p(X) = \Pi_X^{-1}(\max_u \Pi_X(u)).$$

In other words, $V_p(X)$ contains elements of U which are most compatible with the restrictions defined by Π_X . It follows from intuition, and from Luhanjula's formal proof (Luhandjula 1986), that when Π_X is convex, $V_p(X)$ is a real-valued interval, and when Π_X is strongly convex, $V_p(X)$ is a single real number.

Luhandjula considers the mathematical program

$$\begin{aligned} \max \tilde{z} &= \hat{c}x \\ \text{subject to: } \hat{A}_i &\leq \hat{b}_i, \\ x &\geq 0. \end{aligned} \quad (56)$$

By replacing the possibilistic numbers $\hat{c}$, $\hat{A}_i$, and $\hat{b}_i$ with their more possible values, $V_p(\hat{c})$, $V_p(\hat{A}_i)$, and $V_p(\hat{b}_i)$, respectively, Luhandjula arrives at a deterministic equivalent to Eq. (56):

$$\begin{aligned} \max z &= kx \\ \text{subject to: } k_i &\in V_p(\hat{c}_i) \\ \sum_i t_i x_i &\leq s_i \\ t_i &\in V_p(\hat{a}_{ij}) \\ s_i &\in V_p(\hat{b}_i) \\ x &\geq 0. \end{aligned} \quad (57)$$

This formulation varies significantly from the other approaches considered thus far. The possibility of each possibilistic component is maximized individually. Other formulations have required that each possibilistic component $\tilde{c}_j$, $\tilde{a}_{ij}$, and $\tilde{b}_i$ achieve the same possibility level defined by α . This formulation also has a distinct disadvantage over the others we've considered, since to date there is no proposed computational method for determining the "more possible" values, V_p , so there is no way to solve the deterministic MP.

Programming with Fuzzy and Possibilistic Components

Sometimes the values of an optimization problem's components are ambiguous *and* the decision-makers are vague (or flexible) regarding feasibility requirements. This section explores a couple of approaches for dealing with such fuzzy/possibilistic problems.

One type of mixed programming problem that arises (see Dubois and Prade 1980b; Negoita 1981) is a mathematical program with possibilistic constraint coefficients $\tilde{a}_{ij}$ whose possible values are defined by fuzzy numbers of the form $\tilde{a}_{ij}$:

$$\begin{aligned} \max \quad & cx \\ \text{subject to: } & \tilde{a}_i' x' \subseteq \tilde{b}_i \\ & x' = (1, x')t \geq 0. \end{aligned} \quad (58)$$

Zadeh (1965) defines the set-inclusion relation $\tilde{M} \subseteq \tilde{N}$ as $\mu_{\tilde{M}}(r) \leq \mu_{\tilde{N}}(r) \forall r \in R$. Recall that Dubois (1987) interprets the set-inclusive constraint $\tilde{a}_i' x' \subseteq \tilde{b}_i$ as a fuzzy extension of the crisp equality constraint. Mixed programming, however, interprets the set-inclusive constraint to mean that the region in which $\tilde{a}_i' x'$ can possibly occur is restricted to $\tilde{b}_i$, a region which is tolerable to the decision maker. Therefore, the left side of (58) is possibilistic, and the right side is fuzzy.

Negoita (1981) defines the fuzzy right hand side as follows:

$$\tilde{b}_i = \{r | r \geq b_i\}. \quad (59)$$

As a result, we can interpret $\tilde{a}_i' x' \subseteq \tilde{b}_i$ as an extension of an inequality constraint. The set-inclusive constraint (58) is reduced to

$$\begin{aligned} a_i^+(\alpha)x &\leq b_i^+(\alpha) \\ a_i^-(\alpha)x &\geq b_i^-(\alpha) \\ \text{for all } \alpha &\in (0, 1]. \end{aligned} \quad (60)$$

If we abide by Negoita's definition of $\tilde{b}$ (59), $b_i^+ = \infty$ for all values of α , so we can drop the first constraint in (60). Nonetheless, we still have an infinitely (continuum) constrained program,

with two constraints for each value of $\alpha \in (0, 1]$. Inuiguchi observes (Inuiguchi et al. 1990) that if the left-hand side of the membership functions for $a_{i0}, a_{i1}, \dots, a_{in}, b_i$ are identical for all i , and the right-hand side of the membership functions for $a_{i0}, a_{i1}, \dots, a_{in}, b_i$ are identical for all i , the constraints are reduced to the finite set,

$$\begin{aligned} a_{i,\alpha} &\geq b_{i,\alpha} \\ a_{i,\gamma} &\geq b_{i,\gamma} \\ a_{i,\beta} &\leq b_{i,\beta} \\ a_{i,\delta} &\leq b_{i,\delta}. \end{aligned} \quad (61)$$

As per our usual notation, (γ, δ) is the support of the fuzzy number, and (α, β) is its core. In application, constraint formulation (61) has limited utility because of the narrowly defined sets of memberships functions it admits. For example, if $a_{i0}, a_{i1}, \dots, a_{in}, b_i$ are defined by trapezoidal fuzzy numbers, they must all have the same spread, and therefore the same slope, on the right-hand side; and they must all have the same spread, and therefore the same slope, on the left-hand side if (61) is to be implemented. Recall that in this kind of mixed programming, the a_{ij} 's are possibilistic, reflecting a lack of information about their values, and the b_i are fuzzy, reflecting the decision maker's degree of satisfaction with their possible values. It is possible that n possibilistic components and 1 fuzzy component will share identically-shaped distribution, but it is not something likely to happen with great frequency.

Delgado, Verdegay and Vila Delgado, Verdegay, and Villa (1989) propose the following formulation for dealing with ambiguity in the constraint coefficients and right-hand sides, as well as vagueness in the inequality relationship:

$$\begin{aligned} \max \quad & cx \\ \text{subject to: } & \hat{A}x \leq \hat{b} \\ & x \geq 0. \end{aligned} \quad (62)$$

In addition to (62), membership functions $\mu_{a_{ij}}$ are defined for the possible values of each possibilistic element of $\hat{A}$, membership functions μ_{b_i} are defined for the possible values of each

possibilistic element of $\hat{b}$, and membership function μ_i gives the degree to which the fuzzy constraint i is satisfied. Stated another way, μ_i is the membership function of the fuzzy inequality. The uncertainty in the $\tilde{a}_{ij}$ and the $\tilde{b}_i$ is due to ambiguity concerning the actual value of the parameter, while the uncertainty in the $\leq$ is due to the decision maker's flexibility regarding the necessity of satisfying the constraints in full.

Delgado, Verdegay, and Vila do not define the fuzzy inequality, but leave that specification to decision maker. Any ranking index that preserves ranking of fuzzy numbers when multiplied by a positive scalar is allowed. For instance, one could select any of Dubois' four inequalities from section "Fuzzy Relations". Once the ranking index is selected, the problem is solved parametrically, as in Verdegay's earlier work (1982) (see subsection "Verdegay").

To illustrate this approach, let us choose Dubois' pessimistic inequality (i), which interprets $\tilde{A} \geq \tilde{b}$ to mean $\forall x \in A, \forall y \in B, x \geq y$. This is equivalent to $a^+ \geq b^-$. Then (62) becomes

$$\begin{aligned} \max \quad & cx \\ \text{subject to: } & a_{ij}^+(\alpha)x \leq b_i^-(\alpha) \\ & x \geq 0. \end{aligned} \quad (63)$$

Fuzzy/Robust Programming The approach covered so far in this section, has the same α cut define the level of ambiguity in the coefficients and the level at which the decision-maker's requirements are satisfied. These α , however, mean very different things. The fuzzy α represents the level at which the decision-maker's requirements are satisfied. The possibilistic α , on the other hand, represents the likelihood that the parameters will take on values which will attain that level. The solution is interpreted to mean that for any value $\alpha \in (0, 1]$ there is a possibility α of obtaining a solution that satisfies the decision maker to degree α . Using the same α value for both the possibilistic and fuzzy components of the problem is convenient, but does not necessarily provide a meaningful model of reality.

A recent model (Untiedt 2007a) based on Markowitz's mean-variance approach to portfolio

optimization (see Markowitz 1952; Steinbach 2001) differentiates between the fuzzy α and the possibilistic α . Markowitz introduced an efficient combination, which has the minimum risk for a return greater than or equal to a given level; or one which has the maximum return for a risk less than or equal to a given level. The decision maker can move among these efficient combinations, or along the efficient frontier, according to her/his degree of risk aversion.

Similarly, in mixed possibilistic and fuzzy programming, one might wish to allow a trade-off between the potential reward of the outcome and the reliability of the outcome, with the weights of the two competing objectives determined by the decision maker's risk aversion. The desire is to obtain an objective function like the following:

$$\max[\text{reward} + (\text{risk aversion}) \times (\text{reliability})]. \quad (64)$$

The reward variable is the α -level associated with the fuzzy constraints and goal(s). It tells the decision maker how satisfactory the solution is. The reliability variable is the α - level associated with the possibilistic parameters. It tells the decision maker how likely it is that the solution will actually be satisfactory. To avoid confusion, let us refer to the fuzzy constraint membership parameter as α and the possibilistic parameter membership level as β .

In addition, let $\mu \in [0, 1]$ be an indicator of the decision maker's valuation of reward and risk-avoidance, with 0 indicating that the decision maker cares exclusively about the reward, and 1 indicating the only risk avoidance is important. Using this notation, the desired objective is

$$\max(1 - \mu)\alpha + \mu\beta. \quad (65)$$

Suppose we begin with the mixed problem:

$$\begin{aligned} \max \quad & \hat{c}^T x \\ \text{subject to} \quad & \hat{A}x \leq b \\ & x \geq 0. \end{aligned} \quad (66)$$

Incorporating fuzziness from soft constraints in the tradition of Zimmermann and incorporating a pessimistic view of possibility results in the following formulation:

$$\begin{aligned} \max \quad & (1 - \mu)\alpha + \mu\beta \quad (67) \\ \text{subject to} \quad & \alpha \leq -\frac{g}{d_0} + \sum_j \frac{u_j}{d_0} x_j + \sum_j \frac{u_j - w_j}{d_0} x_j \beta \\ & \alpha \leq \frac{b_i}{d_i} - \sum_j \frac{v_{ij}}{d_i} x_j - \sum_j \frac{z_{ij} - v_{ij}}{d_i} x_j \beta \\ & x \geq 0 \\ & \alpha, \beta \in [0, 1]. \end{aligned} \quad (68)$$

The last terms in each of the constraints contain βx , so the system is non-linear. It can be fairly easily reformulated as an optimization program with linear objective function and quadratic constraints, but the feasible set is non-convex, so finding a solution to this mathematical programming problem is very difficult.

Possibilistic, Interval, Cloud, and Probabilistic Optimization Utilizing IVP

This section is taken from (Lodwick and Jamison 2007a, 2008) and begins by defining what is meant by an IVP. This generalization of a probability measure includes probability measures, possibility/necessity measures, intervals, and clouds (see Neumaier 2004) which will allow a mixture of uncertainty within one constraint (in)equality. The previous mixed methods was restricted to a single type of uncertainty for any particular (in)equality and are unable to handle cases in which a mixture of fuzzy and possibilistic parameter occurs in the same constraint (in)equality.

The IVP set function may be thought of as a method for giving a partial representation for an unknown probability measure. Throughout, arithmetic operations involving set functions are in terms of interval arithmetic (Moore 1979) and the set of all intervals contained in $[0, 1]$ is

denoted, $\text{Int}_{[0,1]} = \{[a, b] \mid 0 \leq a \leq b \leq 1\}$. Moreover, S is used to denote the universal set and a set of subsets of the universal set is denoted as $\mathcal{A} \subseteq S$. In particular $\mathcal{A}$ is a set of subset on which a structure has been imposed on it as will be seen and a generic set of the structure $\mathcal{A}$ is denoted by A .

Definition 18 (Weichselberger 2000) Given measurable space $(S, \mathcal{A})$, an interval valued function $i_m : \mathcal{A} \subseteq S \rightarrow \text{Int}_{[0,1]}$ is called an **R-probability** if:

- (a) $i_m(A) = [i_m^-(A), i_m^+(A)] \subseteq [0, 1]$ with $i_m^-(A) \leq i_m^+(A)$,
- (b) $\exists$ a probability measure Pr on $\mathcal{A}$ such that $\forall A \in \mathcal{A}, \text{Pr}(A) \in i_m(A)$. By an **R-probability field** we mean the triple $(S, \mathcal{A}, i_m)$.

Definition 19 (Weichselberger (2000)) Given an R-probability field $\mathcal{R} = (S, \mathcal{A}, i_m)$ the set

$$\mathcal{M}(\mathcal{R}) = \{\text{Pr} \mid \text{Pr} \text{ is a probability measure on } \mathcal{A} \text{ such that } \forall A \in \mathcal{A}, \text{Pr}(A) \in i_m(A)\}$$

is called the **structure** of $\mathcal{R}$.

Definition 20 (Weichselberger 2000) An R-probability field $\mathcal{R} = (S, \mathcal{A}, i_m)$ is called an **F-probability field** if $\forall A \in \mathcal{A}$:

$$i_m^+(A) = \sup\{\text{Pr}(A) \mid \text{Pr} \in \mathcal{M}(\mathcal{R})\},$$

$$i_m^-(A) = \inf\{\text{Pr}(A) \mid \text{Pr} \in \mathcal{M}(\mathcal{R})\}.$$

It is interesting to note that given a measurable space $(S, \mathcal{A})$ and a set of probability measures P , then defining $i_m^+(A) = \sup\{\text{Pr}(A) \mid \text{Pr} \in P\}$ and $i_m^-(A) = \inf\{\text{Pr}(A) \mid \text{Pr} \in P\}$ gives an F-probability and that P is a subset of the structure.

The following examples show how intervals, possibility distributions, clouds and (of course) probability measures can define R-probability fields on $\mathcal{B}$, the Borel sets on the real line.

Example 21 (An Interval Defines an F-Probability Field) Let $I = [a, b]$ be a non-empty interval on the real line. On the Borel sets define

$$i_m^+(A) = \begin{cases} 1 & \text{if } I \cap A \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

and

$$i_m^-(A) = \begin{cases} 1 & \text{if } I \subseteq A \\ 0 & \text{otherwise} \end{cases}$$

then

$$i_m(A) = [i_m^-(A), i_m^+(A)]$$

defines an F-probability field $\mathcal{R} = (R, \mathcal{B}, i_m)$. To see this, simply let P be the set of all probability measures on $\mathcal{B}$ such that $\text{Pr}(I) = 1$.

This example also illustrates that any set A , not just an interval I , can be used to define an F-probability field.

Example 22 (A probability measure is an F-probability field) Let Pr be a probability measure over $(S, \mathcal{A})$. Define

$$i_m(A) = [\text{Pr}(A), \text{Pr}(A)].$$

This definition of a probability as an IVPF is equivalent to having total knowledge about a probability distribution over S . The concept of a cloud was introduced by Neumaier in (Neumaier 2004) as follows:

Definition 23 A **cloud** over set S is a mapping c such that:

1. $\forall s \in S, c(s) = [\underline{n}(s), \bar{p}(s)]$ with $0 \leq \underline{n}(s) \leq \bar{p}(s) \leq 1$
2. $(0, 1) \subseteq \cup_{s \in S} c(s) \subseteq [0, 1]$ In addition, random variable X taking values in S is said to belong to cloud c (written $X \in c$) iff
3. $\forall \alpha \in [0, 1], \text{Pr}(\underline{n}(X) \geq \alpha) \leq 1 - \alpha \leq \text{Pr}(\bar{p}(X) > \alpha)$

Property (3) above defines when a random variable belongs to a cloud. That any cloud contains a random variable X is proved in Sect. 5 of (Neumaier 2005). This is a significant result as will be seen, since among other things, it means that clouds can be used to define IVPs. Clouds are closely related to possibility theory.

It is shown in (Jamison and Lodwick 2002) that possibility distributions can be constructed which satisfy the following consistency definition.

Definition 24 (Jamison and Lodwick 2002) Let $p: S \rightarrow [0, 1]$ be a possibility distribution function with associated possibility measure Pos and necessity measure Nec . Then p is said to be **consistent** with random variable X if $\forall$ measurable sets A , $\text{Nec}(A) \leq \Pr(X \in A) \leq \text{Pos}(A)$.

Possibility distributions constructed in a consistent manner are able to bound (unknown) probabilities of interest. The reason this is significant is twofold. Firstly, possibility and necessity distributions are easier to construct since the axioms they satisfy are more general. Secondly, the algebra on possibility and necessity pairs are much simpler since they are min/max algebras akin to the min/max algebra of interval arithmetic (see, for example, Lodwick and Jamison 2003b). In particular, they avoid convolutions which are requisite for probabilistic arithmetic. The concept of a cloud can be stated in terms of certain pairs of consistent possibility distributions as shown by the following proposition (which means that clouds may be considered as pairs of consistent possibilities – possibility and necessity pairs, see Jamison and Lodwick 2002).

Proposition 25 Let $\bar{p}, \underline{p}$ be a pair of regular possibility distribution functions over set S such that $\forall s \in S \bar{p}(s) + \underline{p}(s) \geq 1$. Then the mapping $c(s) = [\underline{n}(s), \bar{p}(s)]$ where $\underline{n}(s) = 1 - \underline{p}(s)$ (i.e. the dual necessity distribution function) is a cloud. In addition, if X is a random variable taking values in S and the possibility measures associated with $\bar{p}, \underline{p}$ are consistent with X then X belongs to cloud c . Conversely, every cloud

defines such a pair of possibility distribution functions and their associated possibility measures are consistent with every random variable belonging to c .

Proof (see Jamison and Lodwick 2004; Lodwick and Jamison 2006, 2008)

Example 26 (A Cloud Defines an R-Probability Field) Let c be a cloud over the real line. Let $\text{Pos}^1, \text{Nec}^1, \text{Pos}^2, \text{Nec}^2$ be the possibility measures and their dual necessity measures relating to $\bar{p}(s)$ and $\underline{p}(s)$ (where $\bar{p}$ and $\underline{p}$ are as in Proposition 18). Define

$$i_m(A) = [\max\{\text{Nec}^1(A), \text{Nec}^2(A)\}, \min\{\text{Pos}^1(A), \text{Pos}^2(A)\}].$$

Neumaier (2005) proved that every cloud contains a random variable X . Since consistency requires that $\Pr(X \in A) \in i_m(A)$, the result that every cloud contains a random variable X shows consistency. Thus every cloud defines an R-probability field because the inf and sup of the probabilities are bounded by the lower and upper bounds of $i_m(A)$.

Example 27 (A Possibility Distribution Defines an R-Probability Field) Let $p: S \rightarrow [0, 1]$ be a possibility distribution function and let Pos be the associated possibility measure and Nec the dual necessity measure. Define $i_m(A) = [\text{Nec}(A), \text{Pos}(A)]$. Defining a second possibility distribution, $\underline{p}(x) = 1 \forall x$ means that the pair $p, \underline{p}$ define a cloud for which $i_m(A)$ defines the R-probability. Since a cloud defines an R-probability field, this means that this possibility in turn generates a R-probability.

Note that the above example means that every pair of possibility $p(x)$ and necessity $n(x)$ such that $n(x) \leq p(x)$ has an associated F-probability field. This is because such a pair defines a cloud and in every cloud there exists a probability distribution. The F-probability field can then be constructed from a $\inf/\sup$ over all such enclosed probabilities

that are less than or equal to the bounding necessity/possibility distributions.

The application of these concepts to mixed fuzzy and possibilistic optimization is as follows. Suppose the optimization problem is to maximize $f(\vec{x}, \vec{a})$ subject to $g(\vec{x}, \vec{b}) = 0$ (where $\vec{a}$ and $\vec{b}$ are parameters). Assume $\vec{a}$ and $\vec{b}$ are vectors of independent uncertain parameters, each with an associated IVP. Assume the constraint may be violated at a cost $\vec{p} > 0$ so that the problem becomes one to maximize

$$h(\vec{x}, \vec{a}, \vec{b}) = f(\vec{x}, \vec{a}) - \vec{p} | g(\vec{x}, \vec{b}) |.$$

Given the independence assumption, form an IVP for the product space $i_{\vec{a} \times \vec{b}}$ for the joint distribution (see the example below and Jamison and Lodwick 2004; Lodwick and Jamison 2006, 2008) and calculate the interval-valued expected value (see Jamison 1998; Yager 1981) with respect to this IVP. The interval-valued expected value is denoted (there is a lower expected value and an upper expected value)

$$\int_R h(\vec{x}, \vec{a}, \vec{b}) di_{\vec{a} \times \vec{b}}. \quad (69)$$

To optimize (69) requires an ordering of intervals, a *valuation function* denoted by $v: \text{Int}_R \rightarrow \mathbb{R}^n$. One such ordering is the midpoint of the interval on the principle that in the absence of additional data, the midpoint is the best estimate for the true value so that for this function (midpoint and width), $v: \text{Int}_R \rightarrow \mathbb{R}^2$. Thus, for $I = [a, b]$, this particular valuation function is $v(I) = ((a + b)/2, b - a)$. Next a *utility function* of a vector in $\mathbb{R}^n$ is required which is denote by $u: \mathbb{R}^n \rightarrow \mathbb{R}$. A utility function operating on the midpoint and width valuation function is $u: \mathbb{R}^2 \rightarrow \mathbb{R}$ and particular utility is a weighted sum of the midpoint and width $u(c, d) = \alpha c + \beta d$. Using the valuation and utility functions, the optimization problem is:

$$\max_x u \left(v \left(\int_R h(x, a, b) di_{a \times b} \right) \right). \quad (70)$$

Thus there are four steps to obtaining a real-valued optimization problem from an IVP problem. The first step is to obtain interval probability $h(x, a, b)$. The second step is to obtain the IVP expected value $\int_R h(\vec{x}, \vec{a}, \vec{b}) di_{\vec{a} \times \vec{b}}$. The third step is to obtain the vector value of $v: \text{Int}_R \rightarrow \mathbb{R}^n$. The fourth step is to obtain the utility function value of a vector $u: \mathbb{R}^n \rightarrow \mathbb{R}$.

Example 28 Consider the problem

$$\begin{aligned} \max f(x, a) &= 8x_1 + 7x_2 \\ \text{subject to:} \\ g_1(x, b) &= 3x_1 + [1, 3]x_2 - 4 = 0 \\ g_2(x, b) &= \tilde{2}x_1 + 5x_2 - 1 = 0 \\ \vec{x} &\in [0, 2] \end{aligned}$$

where $\tilde{2} = 1/2/3$, that is, $\tilde{2}$ is a triangular possibilistic number with support (Asai and Tanaka 1973; Baudrit et al. 2005) and modal value at 2. For $\vec{p} = (1, 1)^T$,

$$h(x, a, b) = 5x_1 - \tilde{2}x_1 + [3, 5]x_2 - 6$$

so that

$$\begin{aligned} \int_R h(x, a, b) di_{a \times b} &= 5x_1 + \left[\int_0^1 (\alpha - 3)\alpha, \int_0^1 (-1 - \alpha)\alpha \right] x_1 \\ &\quad + [3, 5]x_2 - 6 \\ &= 5x_1 + \left[-\frac{5}{2}, -\frac{3}{2} \right] x_1 + [3, 5]x_2 - 5. \end{aligned}$$

Since the constant -5 will not affect the optimization, it will be removed (then added at the end), so that

$$\begin{aligned} v \left(\int_R h(x, a, b) di_{a \times b} \right) &= v \left(\left[\frac{5}{2}, \frac{7}{2} \right] x_1 + [3, 5]x_2 \right) \\ &= (3, 1)x_1 + (4, 2)x_2. \end{aligned}$$

Let $u(\vec{y}) = \sum_{i=1}^n y_i$ which for the context of this problem yields

$$\begin{aligned}
\max_x z &= u\left(v\left(\int_R h(x, a, b) di_{a \times b}\right)\right) \\
&= \max_{\vec{x} \in [0, 2]} (4x_1 + 6x_2) - 5 \\
&= 20 - 5 = 15 \\
x_1^* &= 2, x_2^* = 2.
\end{aligned}$$

Example 29 (See *Thipwiwatpotjani 2007*)
Consider

$$\begin{aligned}
\max z &= \widehat{2}x_1 - \overline{0}x_2 + [3, 5]x_3 \\
\widehat{4}x_1 + [1, 5]x_2 - 2x_3 - [0, 2] &= 0 \\
6x_1 - \overline{2}x_2 + 9x_3 - 9 &= 0 \\
-2x_1 - [1, 4]x_2 - \widehat{8}x_3 + \overline{5} &= 0 \\
0 \leq x_1 \leq 3, 1 \leq x_2, x_3 \leq 2.
\end{aligned}$$

Here the $\overline{0}$, $\widehat{2}$, and $\overline{5}$ are probability distributions. Note the mixture of three uncertainty types in the third constraint equation. Using the same approach as in the previous example, the optimal values are:

$$\begin{aligned}
z^* &= 3.9179 \\
x_1^* &= 0, \quad x_2^* = 0.4355, \quad x_3^* = 1.0121.
\end{aligned}$$

In (Thipwiwatpotjani 2007) it is shown that these mixed problems arising from linear programs remain linear programs. Thus, the complexity of mixed problems is equivalent to that of linear programming.

Future Directions

The applications of fuzzy optimization seems to be headed toward “industrial strength” problems. Increasingly, each year there are a greater number of applications that appear. Given that greater attention is being given to the semantics of fuzzy optimization and as fuzzy optimization becomes increasingly used in applications, associated algorithms that are more sophisticated, robust, and efficient will need to be developed to handle these more complex problems. It would be interesting to develop modeling languages like GAMS (Brooke et al. 2002), MODLER (Greenberg

1992), or AMPL (Fourer et al. 1993), that support fuzzy data structures. From the theoretical side, the flexibility that fuzzy optimization has with working with uncertainty data that is fuzzy, flexible, and/or possibilistic (or a mixture of these via IVP), means that fuzzy optimization is able to provide an ample approach to optimization under uncertainty. Further research into the development of more robust methods that use fuzzy Banach spaces would certainly provide a deeper theoretical foundation to fuzzy optimization. Fuzzy optimization that utilize fuzzy Banach spaces have the advantage that the problem remains fuzzy throughout and only when one needs to make a decision or implement the solution does one map the solution to a real number (defuzzify). The methods that map fuzzy optimization problems to their real number equivalent defuzzify first and then optimize. Fuzzy optimization problems that optimize in fuzzy Banach spaces keep the solution fuzzy and defuzzify as a last step.

Continued development of clear input and output semantics of fuzzy optimization will greatly aid fuzzy optimization’s applicability and relevance. When fuzzy optimization is used in, for example, an assembly-line scheduling problem and one’s solution is a fuzzy three, how does one convey this solution to the assembly-line manager? Lastly, continued research into handling dependencies in an efficient way would amplify the usefulness and applicability of fuzzy optimization.

Bibliography

- Asai K, Tanaka H (1973) On the fuzzy – mathematical programming, identification and system estimation proceedings of the 3rd IFAC Symposium, The Hague, 12–15 June 1973, pp 1050–1051
- Audin J-P, Frankowska H (1990) Set-valued analysis. Birkhäuser, Boston
- Baudrit C, Dubois D, Fargier H (2005) Propagation of uncertainty involving imprecision and randomness. ISIPTA 2005, Pittsburgh, pp 31–40
- Bector CR, Chandra S (2005) Fuzzy mathematical programming and fuzzy matrix games. Springer, Berlin
- Bellman RE, Zadeh LA (1970) Decision-making in a fuzzy environment. Manag Sci B 17:141–164
- Brooke A, Kendrick D, Meeraus A (2002) GAMS: A User’s guide. Scientific Press, San Francisco (the

- latest updates can be obtained from <http://www.gams.com/>)
- Buckley JJ (1988a) Possibility and necessity in optimization. *Fuzzy Sets Syst* 25(1):1–13
- Buckley JJ (1988b) Possibilistic linear programming with triangular fuzzy numbers. *Fuzzy Sets Syst* 26(1):135–138
- Buckley JJ (1989a) Solving possibilistic linear programming problems. *Fuzzy Sets Syst* 31(3):329–341
- Buckley JJ (1989b) A generalized extension principle. *Fuzzy Sets Syst* 33:241–242
- Delgado M, Verdegay JL, Vila MA (1989) A general model for fuzzy linear programming. *Fuzzy Sets Syst* 29:21–29
- Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) (1994) *Fuzzy optimization: recent advances*. Physica, Heidelberg
- Dempster MAH (1969) Distributions in interval and linear programming. In: Hansen ER (ed) *Topics in interval analysis*. Oxford Press, Oxford, pp 107–127
- Demster AP (1967) Upper and lower probabilities induced by multivalued mapping. *Ann Math Stat* 38:325–339
- Diamond P (1991) Congruence classes of fuzzy sets form a Banach space. *J Math Anal Appl* 162:144–151
- Diamond P, Kloeden P (1994a) *Metric spaces of fuzzy sets*. World Scientific, Singapore
- Diamond P, Kloeden P (1994b) Robust Kuhn–Tucker conditions and optimization under imprecision. In: Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) *Fuzzy optimization: recent advances*. Physica, Heidelberg, pp 61–66
- Dubois D (1987) Linear programming with fuzzy data. In: Bezdek JC (ed) *Analysis of fuzzy information, vol III: applications in engineering and science*. CRC Press, Boca Raton, pp 241–263
- Dubois D, Prade H (1980a) *Fuzzy sets and systems: theory and applications*. Academic Press, New York
- Dubois D, Prade H (1980b) Systems of linear fuzzy constraints. *Fuzzy Sets Syst* 3:37–48
- Dubois D, Prade H (1981) Additions of interactive fuzzy numbers. *IEEE Trans Autom Control* 26(4):926–936
- Dubois D, Prade H (1983) Ranking fuzzy numbers in the setting of possibility theory. *Inf Sci* 30:183–224
- Dubois D, Prade H (1986) New results about properties and semantics of fuzzy set-theoretical operators. In: Wang PP, Chang SK (eds) *Fuzzy sets*. Plenum Press, New York, pp 59–75
- Dubois D, Prade H (1987) Fuzzy numbers: an overview. Tech. Rep. no 219, (LSI, Univ. Paul Sabatier, Toulouse, France). *Mathematics and logic*. In: James Bezdek C (ed) *Analysis of fuzzy information, vol 1, chap 1*. CRC Press, Boca Raton, pp 3–39
- Dubois D, Prade H (1988) *Possibility theory an approach to computerized processing of uncertainty*. Plenum Press, New York
- Dubois D, Prade H (2005) Fuzzy elements in a fuzzy set. *Proceedings of the 11th International Fuzzy System Association World Congress, IFSA 2005, Beijing, July 2005*, pp 55–60
- Dubois D, Moral S, Prade H (1997) Semantics for possibility theory based on likelihoods. *J Math Anal Appl* 205:359–380
- Ertuğrul I, Tuş A (2007) Interactive fuzzy linear programming and an application sample at a textile firm. *Fuzzy Optim Decis Making* 6:29–49
- Fortin J, Dubois D, Fargier H (2008) Gradual numbers and their application to fuzzy interval analysis. *IEEE Trans Fuzzy Syst* 16(2):388–402
- Fourer R, Gay DM, Kernighan BW (1993) *AMPL: a modeling language for mathematical programming*. Scientific Press, San Francisco, CA (the latest updates can be obtained from <http://www.ampl.com/>)
- Fullér R, Keresztfalvi T (1990) On generalization of Nguyen's theorem. *Fuzzy Sets Syst* 41:371–374
- Ghassan K (1982) New utilization of fuzzy optimization method. In: Gupta MM, Sanchez E (eds) *Fuzzy information and decision processes*, North-Holland, pp 239–246
- Gladish B, Parra M, Terol A, Uriá M (2005) Management of surgical waiting lists through a possibilistic linear multiobjective programming problem. *Appl Math Comput* 167:477–495
- Greenberg H (1992) MODLER: modeling by object-driven linear elemental relations. *Ann Oper Res* 38:239–280
- Hanss M (2005) *Applied fuzzy arithmetic*. Springer, Berlin
- Hsu H, Wang W (2001) Possibilistic programming in production planning of assemble-to-order environments. *Fuzzy Sets Syst* 119:59–70
- Inuiguchi M (1992) Stochastic programming problems versus fuzzy mathematical programming problems. *Jpn J Fuzzy Theory Syst* 4(1):97–109
- Inuiguchi M (1997) Fuzzy linear programming: what, why and how? *Tatra Mt Math Publ* 13:123–167
- Inuiguchi M (2007a) Necessity measure optimization in linear programming problems with fuzzy polytopes. *Fuzzy Sets Syst* 158:1882–1891
- Inuiguchi M (2007b) On possibility/fuzzy optimization. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) *Foundations of fuzzy logic and soft computing: 12th International Fuzzy System Association world congress, IFSA 2007, Cancun, June 2007, proceedings*. Springer, Berlin, pp 351–360
- Inuiguchi M, Ramik J (2000) Possibilistic linear programming: a brief review of fuzzy mathematical programming and a comparison with stochastic programming in portfolio selection problem. *Fuzzy Sets Syst* 111:97–110
- Inuiguchi M, Sakawa M (1997) An achievement rate approach to linear programming problems with an interval objective function. *J Oper Res Soc* 48:25–33
- Inuiguchi M, Tanino T (2004) Fuzzy linear programming with interactive uncertain parameters. *Reliab Comput* 10(5):512–527
- Inuiguchi M, Ichihashi H, Tanaka H (1990) Fuzzy programming: a survey of recent developments. In: Slowinski R, Teghem J (eds) *Stochastic versus fuzzy*

- approaches to multiobjective mathematical programming under uncertainty. Kluwer, Dordrecht, pp 45–68
- Inuiguchi M, Ichihashi H, Kume Y (1992a) Relationships between modality constrained programming problems and various fuzzy mathematical programming problems. *Fuzzy Sets Syst* 49:243–259
- Inuiguchi M, Ichihashi H, Tanaka H (1992b) Fuzzy programming: a survey of recent developments. In: Slowinski R, Teghem J (eds) *Stochastic versus fuzzy approaches to multiobjective mathematical programming under uncertainty*. Springer, Berlin, pp 45–68
- Inuiguchi M, Sakawa M, Kume Y (1994) The usefulness of possibilistic programming in production planning problems. *Int J Prod Econ* 33:49–52
- Jamison KD (1998) Modeling uncertainty using probabilistic based possibility theory with applications to optimization. Ph D thesis, University of Colorado Denver, Department of Mathematical Sciences. <http://www-math.cudenver.edu/graduate/thesis/jamison.pdf>
- Jamison KD (2000) Possibilities as cumulative subjective probabilities and a norm on the space of congruence classes of fuzzy numbers motivated by an expected utility functional. *Fuzzy Sets Syst* 111:331–339
- Jamison KD, Lodwick WA (2001) Fuzzy linear programming using penalty method. *Fuzzy Sets Syst* 119: 97–110
- Jamison KD, Lodwick WA (2002) The construction of consistent possibility and necessity measures. *Fuzzy Sets Syst* 132(1):1–10
- Jamison KD, Lodwick WA (2004) Interval-valued probability measures. UCD/CCM Report No. 213, March 2004
- Joubert JW, Luhandjula MK, Ncube O, le Roux G, de Wet F (2007) An optimization model for the management of South African game ranch. *Agric Syst* 92:223–239
- Kacprzyk J, Orlovski SA (eds) (1987a) Optimization models using fuzzy sets and possibility theory. D Reidel, Dordrecht
- Kacprzyk J, Orlovski SA (1987b) Fuzzy optimization and mathematical programming: a brief introduction and survey. In: Kacprzyk J, Orlovski SA (eds) *Optimization models using fuzzy sets and possibility theory*. D Reidel, Dordrecht, pp 50–72
- Kasperski A, Zeilinski P (2007) Using gradual numbers for solving fuzzy-valued combinatorial optimization problems. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) *Foundations of fuzzy logic and soft computing: 12th International Fuzzy System Association world congress, IFSA 2007, Cancun, June 2007, proceedings*. Springer, Berlin, pp 656–665
- Kaufmann A, Gupta MM (1985) *Introduction to fuzzy arithmetic – theory and applications*. Van Nostrand Reinhold, New York
- Kaymak U, Sousa JM (2003) Weighting of constraints in fuzzy optimization. *Constraints* 8:61–78 (also in the 2001 proceedings of IEEE fuzzy systems conference)
- Klir GJ, Yuan B (1995) *Fuzzy sets and fuzzy logic*. Prentice Hall, Upper Saddle River
- Lai Y, Hwang C (1992) *Fuzzy mathematical programming*. Springer, Berlin
- Lin TY (2005) A function theoretic view of fuzzy sets: new extension principle. In: Filev D, Ying H (eds) *Proceedings of NAFIPS05*
- Liu B (1999) *Uncertainty programming*. Wiley, New York
- Liu B (2000) Dependent-chance programming in fuzzy environments. *Fuzzy Sets Syst* 109:97–106
- Liu B (2001) Fuzzy random chance-constrained programming. *IEEE Trans Fuzzy Syst* 9:713–720
- Liu B (2002) *Theory and practice of uncertainty programming*. Physica, Heidelberg
- Liu B, Iwamura K (2001) Fuzzy programming with fuzzy decisions and fuzzy simulation-based genetic algorithm. *Fuzzy Sets Syst* 122:253–262
- Lodwick WA (1990a) Analysis of structure in fuzzy linear programs. *Fuzzy Sets Syst* 38:15–26
- Lodwick WA (1990b) A generalized convex stochastic dominance algorithm. *IMA J Math Appl Bus Ind* 2: 225–246
- Lodwick WA (1999) *Constrained interval arithmetic*. CCM Report 138, Feb. 1999. CCM, Denver
- Lodwick WA (ed) (2004) Special issue on linkages between interval analysis and fuzzy set theory. *Reliab Comput* 10
- Lodwick WA (2007) Interval and fuzzy analysis: a unified approach. In: *Advances in imaging and electronic physics*, vol 148. Elsevier, San Diego, pp 75–192
- Lodwick WA, Bachman KA (2005) Solving large-scale fuzzy and possibilistic optimization problems: theory, algorithms and applications. *Fuzzy Optim Decis Making* 4(4):257–278. (also UCD/CCM Report No. 216, June 2004)
- Lodwick WA, Inuiguchi M (eds) (2007) Special issue on optimization under fuzzy and possibilistic uncertainty. *Fuzzy Sets Syst* 158(1 Sept):17
- Lodwick WA, Jamison KD (1997) A computational method for fuzzy optimization. In: Bilal A, Madan G (eds) *Uncertainty analysis in engineering and sciences: fuzzy logic, statistics, and neural network approach*, chap 19. Kluwer, Norwell
- Lodwick WA, Jamison KD (eds) (2003a) Special issue: interfaces fuzzy Set theory interval analysis. *Fuzzy Sets Syst* 135(1 April):1
- Lodwick WA, Jamison KD (2003b) Estimating and validating the cumulative distribution of a function of random variables: toward the development of distribution arithmetic. *Reliab Comput* 9:127–141
- Lodwick WA, Jamison KD (2005) Theory and semantics for fuzzy and possibilistic optimization, *Proceedings of the 11th International Fuzzy System Association world congress, IFSA 2005, Beijing, July 2005*
- Lodwick WA, Jamison KD (2006) Interval-valued probability in the analysis of problems that contain a mixture of fuzzy, possibilistic and interval uncertainty. In: Demirli K, Akgunduz A (eds) *2006 conference of the North American Fuzzy Information Processing Society*, 3–6 June 2006, Montréal, paper 327137

- Lodwick WA, Jamison KD (2007a) The use of interval-valued probability measures in optimization under uncertainty for problems containing a mixture of fuzzy, possibilistic, and interval uncertainty. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) Foundations of fuzzy logic and soft computing: 12th International Fuzzy System Association world congress, IFSA 2007, Cancun, Mexico, June 2007, proceedings. Springer, Berlin, pp 361–370
- Lodwick WA, Jamison KD (2007b) Theoretical and semantic distinctions of fuzzy, possibilistic, and mixed fuzzy/possibilistic optimization. *Fuzzy Sets Syst* 158(17):1861–1872
- Lodwick WA, Jamison KD (2008) Interval-valued probability in the analysis of problems containing a mixture of fuzzy, possibilistic, probabilistic and interval uncertainty. *Fuzzy Sets Syst*
- Lodwick WA, McCourt S, Newman F, Humphries S (1999) Optimization methods for radiation therapy plans. In: Borgers C, Natterer F (eds) IMA series in applied mathematics – computational radiology and imaging: therapy and diagnosis. Springer, New York, pp 229–250
- Lodwick WA, Neumaier A, Newman F (2001) Optimization under uncertainty: methods and applications in radiation therapy. *Proc 10th IEEE Int Conf Fuzzy Syst* 3:1219–1222
- Luhandjula MK (1986) On possibilistic linear programming. *Fuzzy Sets Syst* 18:15–30
- Luhandjula MK (1989) Fuzzy optimization: an appraisal. *Fuzzy Sets Syst* 30:257–282
- Luhandjula MK (2004) Optimisation under hybrid uncertainty. *Fuzzy Sets Syst* 146:187–203
- Luhandjula MK (2006) Fuzzy stochastic linear programming: survey and future research directions. *Eur J Oper Res* 174:1353–1367
- Luhandjula MK, Ichihashi H, Inuiguchi M (1992) Fuzzy and semi-infinite mathematical programming. *Inf Sci* 61:233–250
- Markowitz H (1952) Portfolio selection. *J Financ* 7:77–91
- Moore RE (1979) Methods and applications of interval analysis. SIAM, Philadelphia
- Negoita CV (1981) The current interest in fuzzy optimization. *Fuzzy Sets Syst* 6:261–269
- Negoita CV, Ralescu DA (1975) Applications of fuzzy sets to systems analysis. Birkhäuser, Boston
- Negoita CV, Sularia M (1976) On fuzzy mathematical programming and tolerances in planning. *Econ Comput Cybern Stud Res* 3(31):3–14
- Neumaier A (2003) Fuzzy modeling in terms of surprise. *Fuzzy Sets Syst* 135(1):21–38
- Neumaier A (2004) Clouds, fuzzy sets and probability intervals. *Reliab Comput* 10:249–272. Springer
- Neumaier A (2005) Structure of clouds. (Submitted – downloadable <http://www.mat.univie.ac.at/~neum/papers.html>)
- Nguyen HT (1978) A note on the extension principle for fuzzy sets. *J Math Anal Appl* 64:369–380
- Ogryczak W, Ruszczyński A (1999) From stochastic dominance to mean-risk models: semideviations as risk measures. *Eur J Oper Res* 116:33–50
- Ralescu D (1977) Inexact solutions for large-scale control problems. In: Proceedings of the 1st international congress on mathematics at the service of man, Barcelona
- Ramík J (1986) Extension principle in fuzzy optimization. *Fuzzy Sets Syst* 19:29–35
- Ramík J, Rimanek J (1985) Inequality relation between fuzzy numbers and its use in fuzzy optimization. *Fuzzy Sets Syst* 16:123–138
- Ramík J, Vlach M (2002a) Fuzzy mathematical programming: a unified approach based on fuzzy relations. *Fuzzy Optim Decis Making* 1:335–346
- Ramík J, Vlach M (2002b) Generalized concavity in fuzzy optimization and decision analysis. Kluwer, Boston
- Riverol C, Pilipovik MV (2007) Optimization of the pyrolysis of ethane using fuzzy programming. *Chem Eng J* 133:133–137
- Riverol C, Pilipovik MV, Carosi C (2007) Assessing the water requirements in refineries using possibilistic programming. *Chem Eng Process* 45:533–537
- Rommelfanger HJ (1994) Some problems of fuzzy optimization with T-norm based extended addition. In: Delgado M, Kacprzyk J, Verdegay J-L, Vila MA (eds) Fuzzy optimization: recent advances. Physica, Heidelberg, pp 158–168
- Rommelfanger HJ (1996) Fuzzy linear programming and applications. *Eur J Oper Res* 92:512–527
- Rommelfanger HJ (2004) The advantages of fuzzy optimization models in practical use. *Fuzzy Optim Decis Making* 3:293–309
- Rommelfanger HJ, Slowinski R (1998) Fuzzy linear programming with single or multiple objective functions. In: Slowinski R (ed) Fuzzy sets in decision analysis, operations research and statistics, The handbooks of fuzzy sets. Kluwer, pp 179–213
- Roubens M (1990) Inequality constraints between fuzzy numbers and their use in mathematical programming. In: Slowinski R, Teghem J (eds) Stochastic versus fuzzy approaches to multiobjective mathematical programming under uncertainty. Kluwer, Dordrecht, pp 321–330
- Russell B (1924) Vagueness. *Aust J Philos* 1:84–92
- Sahinidis N (2004) Optimization under uncertainty: state-of-the-art and opportunities. *Comput Chem Eng* 28: 971–983
- Saito S, Ishii H (1998) Existence criteria for fuzzy optimization problems. In: Takahashi W, Tanaka T (eds) Proceedings of the international conference on nonlinear analysis and convex analysis, Niigata, Japan, 28–31 July 1998. World Scientific Press, Singapore, pp 321–325
- Shafer G (1976) Mathematical a theory of evidence. Princeton University Press, Princeton
- Shafer G (1987) Belief functions and possibility measures. In: James Bezdek C (ed) Analysis of fuzzy information, Mathematics and logic, vol 1. CRC Press, Boca Raton, pp 51–84

- Sousa JM, Kaymak U (2002) Fuzzy decision making in modeling and control. World Scientific Press, Singapore
- Steinbach M (2001) Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Rev* 43(1):31–85
- Tanaka H, Asai K (1984) Fuzzy linear programming problems with fuzzy numbers. *Fuzzy Sets Syst* 13:1–10
- Tanaka H, Okuda T, Asai K (1973) Fuzzy mathematical programming. *Trans Soc Instrum Control Eng* 9(5): 607–613. (in Japanese)
- Tanaka H, Okuda T, Asai K (1974) On fuzzy mathematical programming. *J Cybern* 3(4):37–46
- Tanaka H, Ichihashi H, Asai K (1984) A formulation of fuzzy linear programming problems based on comparison of fuzzy numbers. *Control Cybern* 13(3):185–194
- Tanaka H, Ichihashi H, Asai K (1985) Fuzzy decisions in linear programming with trapezoidal fuzzy parameters. In: Kacprzyk J, Yager R (eds) *Management decision support systems using fuzzy sets and possibility theory*. Springer, Heidelberg, pp 146–159
- Tang J, Wang D, Fung R (2001) Formulation of general possibilistic linear programming problems for complex industrial systems. *Fuzzy Sets Syst* 119:41–48
- Thipwiwatpotjani P (2007) An algorithm for solving optimization problems with interval-valued probability measures. CCM Report, No. 259 (December 2007). CCM, Denver
- Untiedt E (2006) Fuzzy and possibilistic programming techniques in the radiation therapy problem: an implementation-based analysis. Masters thesis, University of Colorado Denver, Department of Mathematical Sciences (July 5, 2006)
- Untiedt E (2007a) A robust model for mixed fuzzy and possibilistic programming. Project report for math 7593: advanced linear programming. Spring, Denver
- Untiedt E (2007b) Using gradual numbers to analyze non-monotonic functions of fuzzy intervals. CCM Report No 258 (December 2007). CCM, Denver
- Untiedt E, Lodwick WA (2007) On selecting an algorithm for fuzzy optimization. In: Melin P, Castillo O, Aguilar LT, Kacprzyk J, Pedrycz W (eds) *Foundations of fuzzy logic and soft computing: 12th International Fuzzy System Association world congress, IFSA 2007, Cancun, Mexico, June 2007, proceedings*. Springer, Berlin, pp 371–380
- Vasant PM, Barsoum NN, Bhattacharya A (2007) Possibilistic optimization in planning decision of construction industry. *Int J Prod Econ* (to appear)
- Verdegay JL (1982) Fuzzy mathematical programming. In: Gupta MM, Sanchez E (eds) *Fuzzy information and decision processes*. North-Holland, Amsterdam, pp 231–237
- Vila MA, Delgado M, Verdegay JL (1989) A general model for fuzzy linear programming. *Fuzzy Sets Syst* 30:21–29
- Wang Z, Klir GJ (1992) Fuzzy measure theory. Plenum Press, New York
- Wang R, Liang T (2005) Applying possibilistic linear programming to aggregate production planning. *Int J Prod Econ* 98:328–341
- Wang S, Zhu S (2002) On fuzzy portfolio selection problems. *Fuzzy Optim Decis Making* 1:361–377
- Weichselberger K (2000) The theory of interval-probability as a unifying concept for uncertainty. *Int J Approx Reason* 24:149–170
- Werners B (1988) Aggregation models in mathematical programming. In: Mitra G (ed) *Mathematical models for decision support*, NATO ASI series, vol F48. Springer, Berlin
- Werners B (1995) Fuzzy linear programming – algorithms and applications. Addendum to the proceedings of ISUMA-NAPIPS'95, College Park, Maryland, 17–20 Sept 1995, pp A7–A12
- Whitmore GA, Findlay MC (1978) *Stochastic dominance: an approach to decision-making under risk*. Lexington Books, Lexington
- Yager RR (1980) On choosing between fuzzy subsets. *Kybernetes* 9:151–154
- Yager RR (1981) A procedure for ordering fuzzy subsets of the unit interval. *Inf Sci* 24:143–161
- Yager RR (1986) A characterization of the extension principle. *Fuzzy Sets Syst* 18:205–217
- Zadeh LA (1965) Fuzzy Sets. *Inf Control* 8:338–353
- Zadeh LA (1968) Probability measures of fuzzy events. *J Math Anal Appl* 23:421–427
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. *Inf Sci, Part I*: 8:199–249; *Part II*: 8:301–357; *Part III*: 9:43–80
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28
- Zeleny M (1994) Fuzziness, knowledge and optimization: new optimality concepts. In: Delgado M, Kacprzyk J, Verdegay J-L, Vila A (eds) *Fuzzy optimization: recent advances*. Physica, Heidelberg, pp 3–20
- Zimmermann HJ (1974) Optimization in fuzzy environment. Paper presented at the International XXI TIMS and 46th ORSA Conference, Puerto Rico Meeting, San Juan, Porto Rico, Oct 1974
- Zimmermann HJ (1976) Description and optimization of fuzzy systems. *Int J Gen Syst* 2(4):209–216
- Zimmermann HJ (1978) Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets Syst* 1:45–55
- Zimmermann HJ (1983) Fuzzy mathematical programming. *Comput Oper Res* 10:291–298

Foundations of Fuzzy Sets Theory

Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Article Outline

Glossary

Definition of the Subject

Introduction

Fuzzy Sets: Basic Definition and Properties

Fuzzy Relations

Linguistic Variable, Fuzzy Conditional Statement, and Compositional Rule of Inference

The Extension Principle

Fuzzy Numbers

Fuzzy Events and Their Probabilities

Defuzzification of Fuzzy Sets

Fuzzy Logic: Basic Issues

Bellman and Zadeh's General Approach to Decision-Making Under Fuzziness

Concluding Remarks

Bibliography

Glossary

Fuzzy Set A mathematical tool that can formally characterize an imprecise concept. Whereas a conventional set to which elements can either belong or not, elements in a fuzzy set can belong (to some extent, from zero, which stands for a full nonbelongingness) to one, which stands for a full belongingness, through all intermediate values.

Fuzzy Relation A mathematical tool that can formally characterize that which is imprecisely specified, notably by using natural language, relations between variables, for instance, *similar*, *much greater than*, *almost equal*, etc.

Extension Principle Makes it possible to extend relations, algorithms, etc., defined from

variables that take on nonfuzzy (e.g., real) values to those that take on fuzzy values.

Linguistic Variable, Fuzzy Conditional Statement, Com-positional Rule of Inference Make it possible to use variables, which take on linguistic (instead of numeric) values to represent relations between such variables, by using fuzzy conditional statements and use them in inference by using the compositional rule of inference.

Fuzzy Event and Its Probability Make it possible to formally define events which are imprecisely specified, like “high temperature,” and calculate their probabilities, for instance, the probability of a “high temperature tomorrow.”

Fuzzy Logic Provides formal means for the representation of, and inference based on imprecisely specified premises and rules of inference; can be understood in different ways, basically as fuzzy logic in a narrow sense, being some type of multivalued logic, and fuzzy logic in a broad sense, being a way to formalize inference based on imprecisely specified premises and rules of inference.

Definition of the Subject

We provide a brief exposition of basic elements of Zadeh's (1965) fuzzy sets theory. We discuss basic properties, operations on fuzzy sets, fuzzy relations and their compositions, linguistic variables, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, fuzzy dynamic systems, etc. We also outline Bellman and Zadeh's (1970) general approach to decision-making in a fuzzy environment which is a point of departure for virtually all fuzzy decision-making, optimization, control, etc., models.

Introduction

This entry is meant to briefly expose a novice reader to basic elements of theory of fuzzy sets

and fuzzy systems viewed for our purposes as an effective and efficient means and calculus to deal with imprecision in the definition of data, information, and knowledge, and to provide tools and techniques for dealing with imprecision therein. Our exposition will be as formal as necessary, of more intuitive and constructive a character, so that fuzzy tools and techniques can be useful for the multidisciplinary audience of this encyclopedia. For the readers requiring or interested in a deeper exposition of fuzzy sets and related concepts, we will recommend many relevant references, mainly books. However, as the number of books and volumes on this topic and its applications in a variety of fields is huge, we will recommend some of them only, mostly those better known ones. For the newest literature entries, the readers should consult the most recent catalogs of major scientific publishers who have books and edited volumes on fuzzy sets/logic and their applications.

Our discussion will proceed, on the other hand, in the *pure* fuzzy setting, and we will not discuss possibility theory (which is related to fuzzy sets theory). The reader interested in possibility theory is referred to, e.g., Dubois and Prade (1985, 1988) or their article in this encyclopedia.

We will consecutively discuss the idea of a fuzzy set, basic properties of fuzzy sets, operations on fuzzy sets, some extensions of the basic concept of a fuzzy set, fuzzy relations and their compositions, linguistic variables, fuzzy conditional statements, and the compositional rule of inference, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, fuzzy dynamic systems, etc. We also outline Bellman and Zadeh's (1970) general approach to decision-making in a fuzzy environment which is a point of departure for virtually all fuzzy decision-making, optimization, control, etc., models.

Fuzzy Sets: Basic Definition and Properties

Fuzzy sets theory, introduced by Zadeh in 1965 (Zadeh 1965), is a simple yet very powerful, effective, and efficient means to represent and handle imprecise information (of vagueness type) exemplified by *tall* buildings, *large*

numbers, etc. We will present fuzzy sets theory as some *calculus of imprecision*, not as a new set theory in the mathematical sense.

The Idea of a Fuzzy Set

From our point of view, the main purpose of a (conventional) set in mathematics is to formally characterize some concept (or property). For instance, the concept of "integer numbers which are greater than or equal three and less than or equal ten" may be uniquely represented just by showing all integer numbers that satisfy this condition, that is, given by the following set: $\{x \in I : 3 \leq x \leq 10\} = \{3, 4, 5, 6, 7, 8, 9, 10\}$ where I is the set of integers. Notice that we need to specify first a *universe of discourse* (universe, universal set, referential, reference set, etc.) that contains all those elements which are relevant for the particular concept as, e.g., the set of integers I in our example.

A conventional set, say A , may be equated with its *characteristic function* defined as

$$\varphi_A : X \rightarrow \{0, 1\} \quad (1)$$

which associates with each element x of a universe of discourse $X = \{x\}$ a number $\varphi(x) \in \{0, 1\}$ such that: $\varphi_A(x) = 0$ means that $x \in X$ does not belong to the set A , and $\varphi_A(x) = 1$ means that x belongs to the set A .

Therefore, for the set verbally defined as integer numbers which are greater than or equal to three and less than or equal to ten, its equivalent set $A = \{3, 4, 5, 6, 7, 8, 9, 10\}$, listing all the respective integer numbers, may be represented by its characteristic function

$$\varphi_A(x) = \begin{cases} 1 & \text{for } x \in \{3, 4, 5, 6, 7, 8, 9, 10\} \\ 0 & \text{otherwise} \end{cases}$$

Notice that in a conventional set there is a clear-cut differentiation between elements belonging to the set and not, i.e., the transition from the belongingness to nonbelongingness is clear-cut and abrupt.

However, it is easy to notice that a serious difficulty arises when we try to formalize by means of a set of vague concepts which are commonly encountered in everyday discourse and

widely used by humans as, e.g., the statement “integer numbers which are *more or less* equal to six.” Evidently, the (conventional) set cannot be used to adequately characterize such an imprecise concept because an abrupt and clear-cut differentiation between the elements belonging and not belonging to the set is artificial here.

This has led Zadeh (1965) to the idea of a *fuzzy set* which is a class of objects with unsharp boundaries, i.e., in which the transition from the belongingness to nonbelongingness is not abrupt; thus, elements of a fuzzy set may belong to it to *partial degrees*, from the full belongingness to the full nonbelongingness through all intermediate values. Notice that this is presumably the most natural and simple way to formally define the imprecision of meaning.

We should therefore start again with a *universe of discourse* (universe, universal set, referential, reference set, etc.) containing all elements relevant for the (imprecise) concept we wish to formally represent. Then, the characteristic function $\varphi : X \rightarrow \{0, 1\}$ is replaced by a *membership function* defined as

$$\mu_A : X \rightarrow [0, 1] \quad (2)$$

such that $\mu_A(x) \in [0, 1]$ is the degree to which an element $x \in X$ belongs to the fuzzy set A : From

$\mu_A(x) = 0$ for the full nonbelongingness to $\mu_A(x) = 1$ for the full belongingness, through all intermediate ($0 < \mu_A(x) < 1$) values.

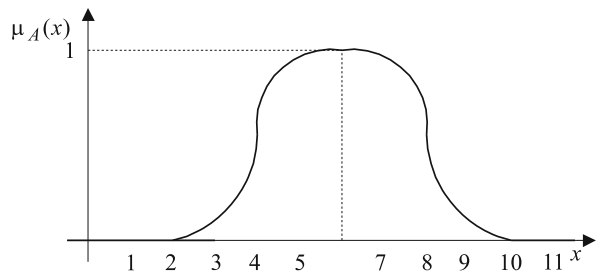
Now, if we consider as an example the concept of integer numbers which are *more or less* six, then $x = 6$ certainly belongs to this set so that $\mu_A(6) = 1$, the numbers five and seven belong to this set *almost surely* so that $\mu_A(5)$ and $\mu_A(7)$ are very close to one, and the more a number differs from six, the less its $\mu_A(\cdot)$. Finally, the numbers below one and above ten do not belong to this set, so that their $\mu_A(\cdot) = 0$. This may be sketched as in Fig. 1 though we should bear in mind that although in our example the membership function is evidently defined for the integer numbers (x 's) only, it is depicted in a continuous form to be more illustrative.

In practice, the membership function is usually assumed to be piecewise linear as shown in Fig. 2 (for the same fuzzy set as in Fig. 1, i.e., the fuzzy set integer numbers which are *more or less* six). To specify the membership function, we then need four numbers only: a, b, c , and d as, e.g., $a = 2$, $b = 5$, $c = 7$, and $d = 10$ in Fig. 2.

Notice that the particular form of a membership function is *subjective* as opposed to an *objective* form of a characteristic function. However, this may be viewed as quite natural as the underlying concepts are subjective indeed as, e.g., the set of integer numbers *more or less* six depend on an

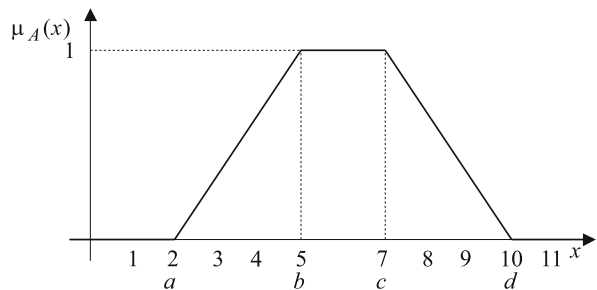
Foundations of Fuzzy Sets Theory,

Fig. 1 Membership function of a fuzzy set, integer numbers which are *more or less* six



Foundations of Fuzzy Sets Theory,

Fig. 2 Membership function of a fuzzy set, integer numbers which are *more or less* six



individual opinion. Unfortunately, this inherent subjectivity of the membership function may lead to some problems in many formal models in which users would rather have a limit to the scope of subjectivity. We will comment on this issue later on.

We now define formally a fuzzy set in a form that is very often used.

A *fuzzy set* A in a universe of discourse $X = \{x\}$, written A in X , is defined as a set of pairs

$$A = \{(\mu_A(x), x)\} \quad (3)$$

where $\mu_A : X \rightarrow [0, 1]$ is the *membership function* of A and $\mu_A(x) \in [0, 1]$ is the *grade of membership* (or a *membership grade*) of an element $x \in X$ in a fuzzy set A .

Needless to say that our definition of a fuzzy set (3) is clearly equivalent to the definition of the membership function (2) because a function may be represented by a set of pairs *argument-value of the function for this argument*. For our purposes, however, the definition (3) is more *set-theoretic-like* which will often be more convenient.

So, in this entry we will practically equate fuzzy sets with their membership functions saying, e.g., a fuzzy set, $\mu_A(x)$, and also very often we will equate fuzzy sets with their labels saying, e.g., a fuzzy set, *large numbers*, with the understanding that the label *large numbers* is equivalent to the fuzzy set mentioned, written $A = \text{large numbers}$. However, we will use the notation $\mu_A(x)$ for the membership function of a fuzzy set A in X , and not an abbreviated notation $A(x)$ as in some more technical papers, to be consistent with our *more-set-theoretic-like* convention.

For practical reasons, it is very often assumed (also in this paper) that all the universes of discourse are finite as, e.g., $X = \{x_1, \dots, x_n\}$. In such a case, the pair $\{(\mu_A(x), x)\}$ will be denoted by $\mu_A(x)/x$ which is called a *fuzzy singleton*.

Then, the fuzzy set A in X will be written as

$$\begin{aligned} A &= \{(\mu_A(x), x)\} = \{\mu_A(x)/x\} = \\ &= \mu_A(x_1)/x_1 + \dots + \mu_A(x_n)/x_n = \sum_{i=1}^n \mu_A(x_i)/x_i \end{aligned} \quad (4)$$

where $+$ and $\sum$ are meant in the set-theoretic sense. By convention, the pairs $\mu_A(x)/x$ with $\mu_A(x) = 0$ are omitted here.

A conventional (nonfuzzy) set may obviously be written in the fuzzy sets notation introduced above, for instance, the (nonfuzzy) set, integer numbers greater than or equal to three and less than or equal to ten, may be written as

$$\begin{aligned} A &= 1/3 + 1/4 + 1/5 + 1/6 + 1/7 + 1/8 \\ &\quad + 1/9 + 1/10 \end{aligned}$$

The family of all fuzzy sets defined in X is denoted by $\mathcal{A}$ it includes evidently also the empty fuzzy set to be defined by (9), i.e., $A = \emptyset$ such that $\mu_A(x) = 0$, for each $x \in X$, and the whole universe of discourse X written as $X = 1/x_1 + \dots + 1/x_n$.

The concept of a fuzzy set as defined above has been the point of departure for the *theory of fuzzy sets* (or *fuzzy sets theory*) which will be briefly sketched below. We will again follow a more intuitive and less formal presentation, which is better suited for this encyclopedia.

Some Extensions of the Concept of Zadeh's Fuzzy Set

The concept of Zadeh's (1965) fuzzy set introduced in the previous section is by far the simplest and most natural way to *fuzzify* the concept of a (conventional) set and clearly provides what we mean to represent and handle imprecision. However, its underlying elements are the most straightforward possible. This concerns above all the membership function. Therefore, it is quite natural that some extensions have been presented to this basic concept. We will just briefly mention some of them.

First, it is quite easy to notice that though the definition of a fuzzy set by the membership function of the type $\mu_A : X \rightarrow [0, 1]$ is the simplest and most straightforward one, allowing for a gradual transition from the belongingness and non-belongingness, it can readily be extended. The same role is namely played by a generalized definition by a membership function of the type

$$\mu_A : X \rightarrow L, \quad (5)$$

where L is some (partially) ordered set as, e.g., a lattice.

This obvious, but powerful extension was introduced by Goguen (1967) as an L -fuzzy set, where L stands for a lattice. Notice that by using a lattice as the set of values of the membership function we can accommodate situations i which we can encounter elements of the universe of discourse which are not comparable.

Another quite obvious extension, already mentioned but not developed by Zadeh (1965, 1983) is the concept of a *type 2 fuzzy set*. The rationale behind this concept is obvious. One can easily imagine that the values of grades of membership of the particular elements of a universe of discourse are fuzzy sets themselves. And further, these fuzzy sets may have grades of membership which are type 2 fuzzy sets, which leads to type 3 fuzzy sets, and one can continue arriving at type n fuzzy sets.

The next, natural extension is that instead of assuming that the degrees of membership are real numbers from the unit interval, one can go a step further and replace these real numbers from $[0,1]$ by intervals with end points belonging to the unit interval. This leads to *interval-valued fuzzy sets* which are attributed to Dubois and Gorzalczany (cf. Klir and Yuan (1995)). Notice that by using intervals as values of degrees of membership, we significantly increase our ability to represent imprecision.

A more radical extension to the concept of Zade's fuzzy set is the so-called *intuitionistic fuzzy set* introduced by Atanassov (1983, 1986).

An intuitionistic fuzzy set A' in a universe of discourse X is defined as

$$A' = \{(x, \mu_{A'}(x), \nu_{A'}(x) \mid x \in X)\} \quad (6)$$

where:

- The degree of membership is

$$\mu_{A'} : X \rightarrow [0, 1],$$

- The degree of nonmembership is

$$\nu_{A'} : X \rightarrow [0, 1],$$

- And the condition holds

$$0 \leq \mu_{A'}(x) + \nu_{A'}(x) \leq 1; \text{ for each } x \in X.$$

Obviously, each (conventional) fuzzy set A in X corresponds to the following intuitionistic fuzzy set A' in X :

$$A = \{x, \mu_A(x), 1 - \mu_A(x) \mid x \in X\} \quad (7)$$

For each intuitionistic fuzzy set A' in X , we call

$$\pi_{A'}(x) = 1 - \mu_{A'}(x) - \nu_{A'}(x); \text{ for each } x \in X \quad (8)$$

the intuitionistic fuzzy index (or a hesitation margin) of x in A' . The intuitionistic fuzzy index expresses a lack of knowledge of whether an element $x \in X$ belongs to an intuitionistic fuzzy set A' or not.

Notice that the concept of an intuitionistic fuzzy set is a substantial departure from the concept of a (conventional) fuzzy set as it assumes that the degrees of membership and non-membership do not sum up to one, as it is the case in virtually all traditional set theories and their extensions. For more information, we refer the reader to Atanassov's (1999) book.

We will not use these extensions in this short introductory entry, and the interested readers are referred to the source literature cited.

Basic Definition and Properties Related to Fuzzy Sets

We will now provide a brief account of basic definitions and properties related to fuzzy sets. We illustrate them with simple examples.

A fuzzy set A is said to be *empty*, written $A = \emptyset$, if and only if

$$\mu_A(x) = 0 \quad \text{for each } x \in X, \quad (9)$$

and since we omit the pairs $0/x$, an empty fuzzy set is really void in the notation (4) as there are no singletons in the right-hand side.

Two fuzzy sets A and B defined in the same universe of discourse X are said to be *equal*, written $A = B$, if and only if

$$\mu_A(x) = \mu_B(x), \quad \text{for each } x \in X \quad (10)$$

Example 1 Suppose that $X = \{1, 2, 3\}$ and

$$A = 0.1/1 + 0.5/2 + 1/3$$

$$B = 0.2/1 + 0.5/2 + 1/3$$

$$C = 0.1/1 + 0.5/2 + 1/3$$

then $A = C$ but $A \neq C$ and $B \neq C$.

It is easy to see that this classic definition of the equality of two fuzzy sets by (10) is rigid and clear-cut, contradicting in a sense our intuitive feeling that the equality of fuzzy sets should be *softer*, and not abrupt, i.e., should rather be to some degree, from zero to one. We will show below one of possible definitions of such an equality to a degree.

Two fuzzy sets A and B defined in X are said to be *equal to a degree* $e(A, B) \in [0, 1]$, written $A =_e B$, and the degree of equality $e(A, B)$ may be defined in many ways exemplified by those given below (cf. Bandler and Kohout (1980)).

First, to simplify, we denote:

Case 1 : $A = B$ in the sense of (10);

Case 2 : $A \neq B$ in the sense of (10) and $T = \{x \in X : \mu_A(x) \neq \mu_B(x)\}$;

Case 3 : $A \neq B$ in the sense of (10) and there exists an $x \in X$ such that

$$\mu_A(x) = 0 \text{ and } \mu_B(x) \neq 0 \quad \text{or} \quad \mu_A(x) \neq 0 \text{ and } \mu_B(x) = 0.$$

Case 4 : $A \neq B$ in the sense of (10) and there exists an $x \in X$ such that

$$\mu_A(x) = 0 \text{ and } \mu_B(x) = 1 \quad \text{or} \quad \mu_A(x) = 1 \text{ and } \mu_B(x) = 0.$$

Now, the following degrees of equality of two fuzzy sets, A and B , may be defined:

$$e_1(A, B) = \begin{cases} 1 & \text{for case 1} \\ \bigwedge_{x \in T} [\mu_A(x) \wedge \mu_B(x)] & \text{for case 2} \\ 0 & \text{for case 3} \end{cases} \quad (11)$$

$$e_2(A, B) = \begin{cases} 1 & \text{for case 1} \\ \bigwedge_{x \in T} [\mu_A(x)/\mu_B(x) - \mu_B(x)/\mu_A(x)] & \text{for case 2} \\ 0 & \text{for case 3} \end{cases} \quad (12)$$

$$e_3(A, B) = \begin{cases} 1 & \text{for case 1} \\ 1 - \max_{x \in X} \mu_A(x) - \mu_B(x) & \text{for case 2} \\ 0 & \text{for case 4} \end{cases} \quad (13)$$

$$e_4(A, B) = \begin{cases} 1 & \text{for case 1} \\ \max_{x \in X} \{[(1 - \mu_A(x)) \wedge [\mu_A(x) \vee (1 - \mu_B(x))]]\} & \text{for case 2} \\ 0 & \text{for case 4} \end{cases} \quad (14)$$

Now we will proceed to the second basic concept of the containment between two fuzzy sets.

A fuzzy set A defined in X is said to be *contained in* or, alternatively, is said to be a *subset* of a fuzzy set B in X , written $A \subseteq B$, if and only if

$$\mu_A(x) \leq \mu_B(x), \quad \text{for each } x \in X \quad (15)$$

Example 2 Suppose that $X = \{1, 2, 3\}$ and

$$\begin{aligned} A &= 0.1/1 + 0.5/2 + 1/4 \\ B &= 0.1/1 + 0.4/2 + 0.9/3 \\ C &= 0.1/1 + 0.6/2 + 1/3, \end{aligned}$$

then only $B \subseteq A$.

This traditional definition of containment is clearly rigid and clear-cut, and hence there have been proposed many other *softer* definitions in which a *degree of containment*, $c(A, B) \in [0, 1]$, has been employed. Once again, Bandler and Kohout's (1980) definitions can be mentioned here, and these basically follow the line of reasoning analogous to that behind the degree of equality, i.e., (11, 12, 13, and 14). The above definitions of the degree of equality and containment are popular but not the only possible ones; some remarks can be found in the books by Dubois and Prade (1980)!, Klir and Folger (1988), and Klir and Yuan (1995).

Let us proceed now to some further relevant foundational notions.

A fuzzy set A defined in X is said to be *normal* if and only if

$$\max_{x \in X} \mu_A(x) = 1 \quad (16)$$

i.e., when the membership function takes on the value of one for at least one argument. Otherwise, the fuzzy set is said to be *subnormal*.

Example 3 If $X = \{1, 2, 3\}$, $A = 0.1/1 + 0.5/2 + 1/3$ and $B = 0.1/1 + 0.6/2 + 0.9/3$, then A is normal and B is subnormal.

Normally, it is desirable to work with normal fuzzy sets since they may provide for some sort of *context-free* comparability, or a common ground or denominator. However, in many instances we obtain in the course of algorithms or procedures subnormal fuzzy sets. They are then often normalized although, unfortunately, the normalization is not a straightforward solution and should be applied with care after some consideration.

We have now some important concepts of non-fuzzy sets associated with a fuzzy set.

The *support* of a fuzzy set A in X , written $\text{supp } A$, is the following (nonfuzzy) set

$$\text{supp } A = \{x \in X : \mu_A(x) > 0\} \quad (17)$$

and, evidently, $\emptyset \subseteq \text{supp } A \subseteq X$.

Example 4 If $X = \{1, 2, \dots, 7\}$ and $A = 0.1/3 + 0.5/4 + 0.8/5 + 1/6$, then $A = \{3, 4, 5, 6\} \subset \{1, 2, \dots, 7\}$.

The α -cut, or α -level set, of a fuzzy set A in X , written A_α , is defined as the following (nonfuzzy) set

$$A_\alpha = \{x \in X : \mu_A(x) \geq \alpha\}, \quad \text{for each } \alpha \in (0, 1] \quad (18)$$

and if $\geq$ in (18) is replaced by $>$, then we have the *strong α -cut*, or *strong α -level set*, of a fuzzy set A in X . In principle, we will use the α -cuts given by (18) if not otherwise specified.

Example 5 If $X = \{1, 2, 3, 4\}$ and $A = 0.1/1 + 0.5/2 + 0.8/3 + 1/4$, then we obtain the following α -cuts

$$\begin{aligned} A_{0.1} &= \{1, 2, 3, 4\} & A_{0.5} &= \{2, 3, 4\} & A_{0.8} \\ &= \{3, 4\} & A_1 &= \{4\}. \end{aligned}$$

The α -cuts have many interesting and relevant properties, and among them one can mention the following one

$$\alpha_1 \leq \alpha_2 \Leftrightarrow A_{\alpha_1} \subseteq A_{\alpha_2} \quad (19)$$

The α -cuts play an extremely relevant role in both formal analysis and applications as they make it possible to uniquely replace a fuzzy set by a sequence of nonfuzzy sets. We will widely use them in the sequel, and the interested reader is referred for details and properties to any book on fuzzy sets theory, e.g., Dubois and Prade (1980), Klir and Folger (1988), or Klir and Yuan (1995).

The following theorem, called the *representation theorem* (cf. Negoita and Ralescu (1975)), is very relevant both in theoretical analysis and applications.

Theorem 1 Each fuzzy set A in X can be represented as

$$A = \sum_{\alpha \in (0, 1]} \alpha A_\alpha, \quad (20)$$

where A_α is an α -cut of A defined as (19), Σ is in the set-theoretic sense, and αA_α denotes the fuzzy set whose degrees of membership are

$$\mu_{\alpha A_\alpha}(x) = \begin{cases} \alpha & \text{for } x \in A_\alpha \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

The expression (20) is also called the *resolution identity*.

Example 6 Let $X = \{1, 2, \dots, 10\}$, and $A = 0.1/2 + 0.3/3 + 0.6/4 + 0.8/5 + 1/6 + 0.7/7 + 0.4/8 + 0.2/9$. Then:

$$\begin{aligned} A &= \sum_{\alpha \in (0, 1]} \alpha A_\alpha \\ &= 0.1(1/2 + 1/3 + 1/4 + 1/5 + 1/6 + 1/7 \\ &\quad + 1/8 + 1/9) + 0.3(1/3 + 1/4 + 1/5 + 1/6 \\ &\quad + 1/7 + 1/8 + 1/9) + 0.6(1/4 + 1/5 + 1/6 + 1/7) \\ &\quad + 0.7(1/5 + 1/6 + 1/7) \\ &\quad + 0.8(1/5 + 1/6) + 1(1/6) \\ &= 0.1/2 + 0.3/3 + 0.6/4 + 0.8/5 + 1/6 \\ &\quad + 0.7/7 + 0.4/8 + 0.2/9. \end{aligned}$$

Notice that the very essence of the presentation theorem is that each fuzzy set can be uniquely represented by a set of its α -cuts.

An important issue, both in theory and application, is to be able to define the *cardinality* of a fuzzy set, i.e., to define how many elements it contains. Unfortunately, this is a difficult problem, and the definitions proposed have been criticized. We will discuss below two of them which are presumably the most widely used.

A *nonfuzzy cardinality* of a fuzzy set $A = \mu_A(x_1)/x_1 + \dots + \mu_A(x_n)/x_n$, the so-called *sigma-count*, denoted $\Sigma\text{Count}(A)$, is defined as (cf. Zadeh (1978, 1983))

$$\Sigma\text{Count}(A) = \sum_{i=1}^n \mu_A(x_i) \quad (22)$$

Example 7 If $A = 1/x_1 + 0.8/x_2 + 0.6/x_3 + 0.2/x_4 + 0/x_5$, then

$$\Sigma\text{Count}(A) = 1 + 0.8 + 0.6 + 0.2 = 2.6.$$

The ΣCount is very simple and is hence widely used. However, an immediate objection may be that the set is fuzzy, but its cardinality is not. A solution in this respect, a *fuzzy cardinality*, was proposed by Zadeh (1983), and it is shown below. Unfortunately, it is more complicated than a nonfuzzy cardinality defined by (22).

Let A be a fuzzy set defined in X , and A_α ; for each $\alpha \in (0, 1]$, its α -cuts are defined by (19). First, Zadeh (1983) introduces the $FG\text{Count}(A)$ as the fuzzy integer defined as

$$FG\text{Count}(A) = \{1/0\} \sum_{\alpha \in (0, 1]} \alpha / \text{card}(A_\alpha) \quad (23)$$

where $\sum$ is in the set-theoretic [cf. (Bandemer and Gottwald 1995)], (A_α) is the usual number of elements in A_α , and $1/0$ means the integer number 0.

Equivalently, if A is defined in X such that $\text{card}(X) = n$, then for each nonnegative integer $i = 0, 1, \dots, n$, we denote

$$FG\text{Count}(A)_i = \sum_{\alpha \in (0, 1]} \{\alpha : \text{card}(A_\alpha) \geq i\}. \quad (24)$$

Semantically, $FG\text{Count}(A)_i$ is the truth of the proposition A contains at least i elements.

Next, Zadeh (1983) introduces the $FL\text{Count}(A)$ which is defined as the truth of the proposition, “ A contains at most i elements,” i.e.,

$$FL\text{Count}(A) = \neg[FG\text{Count}(A)] - 1, \quad (25)$$

where $\neg[\cdot]$ is the complement to be defined by (32) and $-$ is the subtraction in the sense of fuzzy numbers (64).

Similarly as in (24), if A is defined in $X = \{1, 2, \dots, n\}$, then we can denote

$$FL\text{Count}(A)_i = \sup_{\alpha \in (0, 1]} \{\alpha : \text{card}(A_\alpha) \geq n - i\} \quad (26)$$

Notice that

$$FG\text{Count}(A)_i = 1 - FL\text{Count}(A)_{i+1} \text{ for } i = 1, 2, \dots, n \quad (27)$$

Finally, Zadeh (1983) introduces the $FEC\text{Count}(A)$ as

$$FEC\text{Count}(A) = FG\text{Count}(A) \cap FL\text{Count}(A) \quad (28)$$

or, similarly as above,

$$FEC\text{Count}(A)_i = FG\text{Count}(A)_i \wedge FL\text{Count}(A)_i, \quad i = 1, 2, \dots, n \quad (29)$$

where $\cap$ and $\wedge$ denote the intersection of two fuzzy sets and the minimum operations, respectively, as in (34).

Example 8 For the same fuzzy set as in Example 7, i.e., $A = 1/x_1 + 0.8/x_2 + 0.6/x_3 + 0.2/x_4 + 0/x_5$, we obtain

$$FG\text{Count}(A) = 1/0 + 1/1 + 0.8/2 + 0.6/3 + 0.2/4 + 0/5$$

$$FL\text{Count}(A) = 0/0 + 0.2/1 + 0.4/2 + 0.6/3 + 0.2/4 + 0/5$$

$$FEC\text{Count}(A) = 0/0 + 0.2/1 + 0.4/2 + 0.6/3 + 0.2/4 + 0/5.$$

The above classical definitions of the cardinality of a fuzzy set are widely employed, in particular the nonfuzzy cardinality ΣCount . However, the problem of how to define the cardinality of a

fuzzy set is conceptually difficult, and the best source are here Wygalak's (1996, 2003) books.

An important issue, which is widely used in applications, is a *distance* between two fuzzy sets. In practice, normalized distances are clearly more interesting. In the literature (Kacprzyk 1983), and in this book too, the following two basic definitions are used.

Suppose that we have two fuzzy sets, A and B , both defined in $X = \{x_1, \dots, x_n\}$. Then, we have the following two basic (normalized) distances:

- The *normalized linear* (Hamming) *distance* between A and B in X is defined as

$$l(A, B) = \frac{1}{n} \sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|. \quad (30)$$

- The *normalized quadratic* (Euclidean) *distance* between A and B in X is defined as

$$q(A, B) = \sqrt{\frac{1}{n} \sum_{i=1}^n [\mu_A(x_i) - \mu_B(x_i)]^2} \quad (31)$$

Example 9 If $X = \{1, 2, \dots, 7\}$, $A = 0.7/1 + 0.2/2 + 0.6/4 + 0.5/5 + 1/6$ and $B = 0.2/1 + 0.6/4 + 0.8/5 + 1/7$, then:

$$l(A, B) = 0.37 \quad q(A, B) = 0.49$$

Now we will proceed to the basic set-theoretic and algebraic operations on fuzzy sets. They are clearly crucial for both theoretical analysis and applications.

Basic Operations on Fuzzy Sets

Similarly as in the conventional (nonfuzzy) set theory, the basic operations in fuzzy set theory are also the complement, intersection, and union which will be defined below.

The *complement* of a fuzzy set A in X , written $\neg A$, is defined as

$$\mu_{\neg A}(x) = 1 - \mu_A(x), \quad \text{for each } x \in X, \quad (32)$$

and the complement corresponds to the negation, *not*.

Example 10 If $X = \{1, 2, 3\}$ and $A = 0.1/1 + 0.7/2 + 1/3$, then $\neg A = 0.9/1 + 0.3/2$.

The idea of the complement can be portrayed as in Fig. 3.

This definition is the most widely used due to its simplicity and some important mathematical properties. Sometimes different definitions can be justified and useful in specific contexts, for instance, when $X = [0, 1]$, we have the following

$$\mu_{\neg A}(x) = \mu_A(1 - x), \quad \text{for each } x \in [0, 1] \quad (33)$$

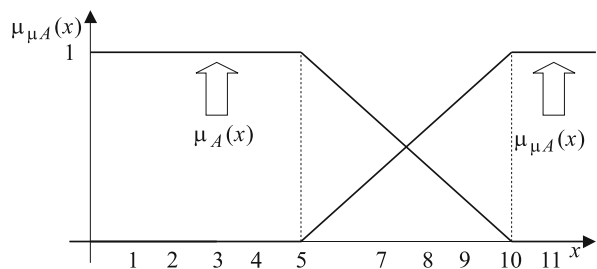
The *intersection* of two fuzzy sets A and B in X , written $A \cap B$, is defined as

$$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x), \quad \text{for each } x \in X \quad (34)$$

where $\wedge$ is the minimum operation, i.e., $a \wedge b = \min(a, b)$; the intersection of two fuzzy sets corresponds to the connective *and*

Example 11 If $X = \{1, 2, 3, 4\}$, and $A = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$ and $B = 1/1 + 0.8/2 + 0.5/3 +$

Foundations of Fuzzy Sets Theory, Fig. 3 The complement of a fuzzy set



0.2/4, then we obtain by (34) $A \cap B = 0.2/1 + 0.5/2 + 0.5/3 + 0.2/4$.

The intersection can be illustrated as in Fig. 4 where $\mu_{A \cap B}(x)$ is shown in bold line.

The union of two fuzzy sets A and B in X , written $A + B$, is defined as

$$\mu_{A+B}(x) = \mu_A(x) \vee \mu_B(x), \quad \text{for each } x \in X \quad (35)$$

where $\vee$ is the maximum operation, i.e., $a \vee b = \max(a, b)$; the union of two fuzzy sets corresponds to the connective “or.”

Example 12 If $X = \{1, 2, 3, 4\}$, and $A = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$ and $B = 1/1 + 0.8/2 + 0.5/3 + 0.2/4$, then $A + B = 1/1 + 0.8/2 + 0.8/3 + 1/4$.

The union can be portrayed as in Fig. 5 in which $\mu_{A+B}(x)$ is shown in bold line.

The above definitions of the basic operations are widely used, and justified. However, other definitions are often employed too. In particular, for the intersection and union the t -norms and s -norms are popular.

A t -norm is defined as

$$t : [0, 1] \times [0, 1] \rightarrow [0, 1] \quad (36)$$

such that, for each $a, b, c \in [0, 1]$:

1. It has 1 as the unit element, i.e., $t(a, 1) = a$,
2. It is monotone, i.e., $a \leq b \Rightarrow t(a, c) \leq t(b, c)$,
3. It is commutative, i.e., $t(a, b) = t(b, a)$, and
4. It is associative, i.e., $t[a, t(b, c)] = t[t(a, b), c]$.

Some more relevant examples of t -norms are the following:

- The minimum (which is the most widely used)

$$t(a, b) = a \wedge b = \min(a, b) \quad (37)$$

- The algebraic product

$$t(a, b) = a \cdot b \quad (38)$$

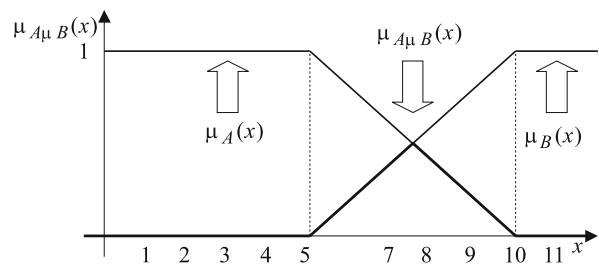
- The Łukasiewicz t -norm

$$t(a, b) = \max(0, a + b - 1), \quad (39)$$

and notice that we have written above both $t(a, b)$ and atb .

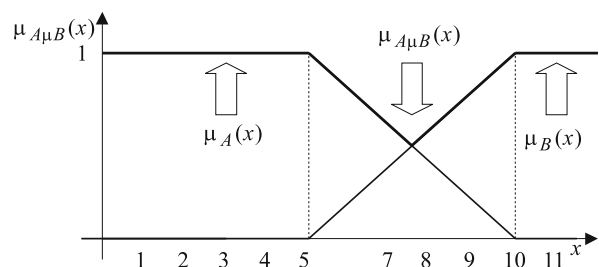
Foundations of Fuzzy Sets Theory,

Fig. 4 Membership function of a fuzzy set, integer numbers which are more or less six



Foundations of Fuzzy Sets Theory, Fig. 5

The union of two fuzzy sets



An *s*-norm (or a *t*-conform) is defined as

$$s : [0, 1] \times [0, 1] \rightarrow [0, 1] \quad (40)$$

such that, for each $a, b, c \in [0, 1]$:

1. It has 0 as the unit element, i.e., $s(a, 0) = a$.
2. It is monotone, i.e., $a \leq b \Rightarrow s(a, c) \leq s(b, c)$.
3. It is commutative, i.e., $s(a, b) = s(b, a)$.
4. It is associative, i.e., $s[a, s(b, c)] = s[s(a, b), c]$.

Some more relevant examples of *s*-norms are the following:

- The maximum (which is the most widely used)

$$s(a, b) = a \vee b = \max(a, b) \quad (41)$$

- The probabilistic product

$$s(a, b) = a + b - ab \quad (42)$$

- The Łukasiewicz *s*-norm

$$s(a, b) = \min(a + b, 1) \quad (43)$$

Notice that a *t*-norms is *dual* to an *s*-norms in that $s(a, b) = 1 - t(1 - a, 1 - b)$.

The *t*-norms and *s*-norms are very important for fuzzy sets theory, and among a multitude of works dealing with them, the most complete reference is provided by Klement, Mesiar, and Pap's (2000) book.

Notice that while defining the basic concepts and operations on fuzzy sets, we have always referred to their linguistic sense. This is very important because without such a linguistic connection, it would have been very difficult to provide semantics of both the very concept of a fuzzy set and the operations. This is very relevant and has great implications for the development of fuzzy sets-related tools and techniques as we will see later.

As to some other operations on fuzzy sets that may be of use in this book, one should also mention the following ones.

The *product of a scalar* $a \in R$ and a fuzzy set A in X , written aA , is defined as

$$\mu_{aA}(x) = a\mu_A(x), \quad \text{for each } x \in X \quad (44)$$

where, by necessity, $0 \leq a \leq 1/\mu_A(x)$, for each $x \in X$.

The *k*th power of a fuzzy set A in X , written A^k , is defined as

$$\mu_{A^k}(x) = [\mu_A(x)]^k, \quad \text{for each } x \in X, \quad (45)$$

where $k \in R$ and, evidently, $0 \leq [\mu_A(x)]^k \leq 1$.

An important issue is the *adequacy* of the operations on fuzzy sets, i.e., whether they do reflect the real human perception of their essence, i.e., whether they really reflect the semantics of “not,” “and,” “or,” etc. Diverse approaches have been used to find and justify a particular definition. These approaches may be classified as:

- *Intuitive* as the original Zadeh's (1965, 1973) works in which it is shown by a rational argument that the operations defined are proper.
- *Axiomatic* whose line of reasoning is to assume some set of plausible conditions to be fulfilled, and then to show using some analytic tools that definitions assumed are the only possible ones; this may be exemplified by Bellman and Giertz (1973).
- *Experimental* whose essence is to devise some psychological tests for a group of certain individuals, and then use the responses to find which operation is best justified; this may be exemplified by Zimmermann and Zysno (1980).

Now we will present the concept of a fuzzy relation which is, as its nonfuzzy counterpart, crucial for the theory and applications.

Fuzzy Relations

The concept of a relation is crucial for virtually all areas of mathematics and its applications, and the same holds true for fuzzy sets theory and its applications.

A *fuzzy relation* R between two (nonfuzzy) sets $X = \{x\}$ and $Y = \{y\}$ is defined as a fuzzy set in the Cartesian product $X \times Y$, i.e.,

$$R = \{(\mu_R(x, y), (x, y))\} \\ = \{\mu_R(x, y)/(x, y)\} \text{ for each } (x, y) \in X \times Y \quad (46)$$

where $\mu_R(x, y) : X \times Y \rightarrow [0, 1]$ is the membership function of the fuzzy relation R , and $\mu_R(x, y) \in [0, 1]$ gives the degree to which the elements $x \in X$ and $y \in Y$ are between each other in relation to R .

The above fuzzy relation is defined in the Cartesian product of (nonfuzzy!) two sets, X and Y , and is called a binary fuzzy relation. In general, a fuzzy relation may be defined in the Cartesian product of k sets, $X_1 \times \dots \times X_k$, and is then called a k -ary fuzzy relation. In this perspective, a fuzzy set is an unary fuzzy relation.

Example 13 If $X = \{\text{horse, donkey}\}$ and $Y = \{\text{mule, cow}\}$, then the fuzzy relation R labeled *similarity* may be exemplified by

$$R = \text{similarity} \\ = 0.8/(\text{horse, mule}) + 0.4/(\text{horse, cow}) \\ + 0.9/(\text{donkey, mule}) \\ + 0.2/(\text{donkey, cow})$$

to be read as: The horse and the mule are similar (with respect to *our own* subjective aspects!) to degree 0.8, i.e., to a very high extent, the horse and the cow are similar to degree 0.4, i.e., to quite a low extent, etc.

It may easily be seen that the concept of a fuzzy relation makes it possible to express an imprecise, or imprecisely specified, relationship between elements of some sets, as opposed to a precise and abrupt one in the case of a nonfuzzy relation in which any two elements can either be or not in relation.

A fuzzy relation R in $X \times Y$ for X and Y of a sufficiently low dimensionality may be conveniently represented in matrix form exemplified, for the fuzzy relation $R = \text{similarity}$ in Example 13, by

$$R = \text{“similarity”} = \begin{array}{c|cc} & y = \text{mule} & \text{cow} \\ \hline x = \text{horse} & 0.8 & 0.4 \\ \text{donkey} & 0.9 & 0.2 \end{array}$$

Since a fuzzy relation is a fuzzy set, all the definitions, properties, operations, etc. on fuzzy sets presented in sections “[The Idea of a Fuzzy Set](#)” and “[Basic Operations on Fuzzy Sets](#)” hold as well, and we will concentrate below on those more specific ones.

The *max-min composition* of two fuzzy relations R in $X \times Y$ and S in $Y \times Z$, written $R \circ_{\max-\min} S$, is defined as a fuzzy relation in $X \times Z$ such that

$$\mu_{R \circ_{\max-\min} S}(x, z) = \max_{y \in Y} [\mu_R(x, y) \wedge \mu_S(y, z)] \quad (47)$$

for each $x \in X, z \in Z$

and since this type of composition will be used throughout this entry, if not otherwise specified, then it will be briefly denoted as $R \circ S$.

Example 14 If $X = \{1, 2\}$, $Y = \{1, 2, 3\}$ and $Z = \{1, 2, 3, 4\}$, and the fuzzy relations R and S are as below, its resulting max-min composition, $R \circ S$, is then:

$$R \circ S =$$

$$= \begin{array}{c|ccc} & y = 1 & 2 & 3 \\ \hline x = 1 & 0.3 & 0.8 & 1 \\ 2 & 0.9 & 0.7 & 0.4 \end{array}$$

$$\circ \begin{array}{c|cccc} & z = 1 & 2 & 3 & 4 \\ \hline y = 1 & 0.7 & 0.6 & 0.4 & 0.1 \\ 2 & 0.4 & 1 & 0.7 & 0.2 \\ 3 & 0.5 & 0.9 & 0.6 & 0.8 \end{array} =$$

$$= \begin{array}{c|cccc} & z = 1 & 2 & 3 & 4 \\ \hline x = 1 & 0.5 & 0.9 & 0.7 & 0.8 \\ 2 & 0.7 & 0.7 & 0.7 & 0.4 \end{array}$$

This max-min composition of fuzzy relations is the original Zadeh’s definition (cf. Zadeh (1973)) and is certainly the most widely used. However, if we notice that the two basic operations involved in the definition of their composition, i.e., *min* ($\wedge$)

and $\max(\vee)$, are just specific examples of the t -norm and s -norm (t -conorm) discussed in section “[Basic Operations on Fuzzy Sets](#),” then one can well define a much more general type of composition given below.

The s - t -norm composition of two fuzzy relations R in $X \times Y$ and S in $Y \times Z$, written $R \circ_{s-t}$, is defined as a fuzzy relation in $X \times Z$ such that

$$\mu_{R \circ_{s-t} S}(x, z) = s_{y \in Y} [\mu_R(x, y) t \mu_S(y, z)] \quad (48)$$

for each $x \in X, \quad z \in Z$.

Fuzzy relations, similar to their nonfuzzy counterparts, play a crucial role in virtually all aspects of the theory and application of fuzzy sets, notably in rule-based fuzzy modeling discussed later in this entry. An important issue is related to so-called fuzzy relational equations. We will not deal with this due to lack of space, and we refer the interested reader to, for instance, the recent book by Peeva and Kyosev (2005).

Finally, let us mention two concepts concerning the fuzzy sets that are related to fuzzy relations.

The *Cartesian product* of two fuzzy sets A in X and B in Y , written $A \times B$, is defined as a fuzzy set in $X \times Y$ such that

$$\mu_{A \times B}(x, y) = [\mu_A(x) \wedge \mu_B(y)] \text{ for each } x \in X, y \in Y. \quad (49)$$

A fuzzy relation R in $X \times Y \times \dots \times Z$ is said to be *decomposable* if and only if it can be represented as

$$\mu_R(x, y, \dots, z) = \mu_{R_x}(x) \wedge \mu_{R_y}(y) \wedge \dots \wedge \mu_{R_z}(z) \quad (50)$$

for each $x \in X, \quad y \in Y, \dots, \quad z \in Z$

where $\mu_{R_x}(x), \mu_{R_y}(y), \dots, \mu_{R_z}(z)$ are projections of the fuzzy relation $\mu_R(x, y)$ on $X, Y, \dots, Z$, respectively, defined as

$$\mu_{R_x}(x) = \sup_{\{y, \dots, z\} \in Y \times Z} \mu_R(x, y, \dots, z) \text{ for each } x \in X \quad (51)$$

Fuzzy relations play a major role in fuzzy sets theory and its applications and will also be relevant for some of our next considerations.

Linguistic Variable, Fuzzy Conditional Statement, and Compositional Rule of Inference

We have already mentioned that while defining various concepts and properties of fuzzy sets, it is expedient to use natural language because linguistic terms and descriptions best provide semantics. Now we will briefly present the essence of Zadeh's (1973) *linguistic approach* in which this fact has been exploited to an even greater extent. This approach has inspired and triggered many new areas of research, notably fuzzy control which has resulted in so many real-world applications in diverse areas and has been a decisive factor in the so-called *fuzzy boom* which was started in the mid-1980s by the launching of a multitude of domestic appliances and professional products exemplified by fuzzy logic-controlled washing machines, cameras, automobile automatic transmissions, cranes, subway trains, etc. These applications have been decisive for a wide acceptance of fuzzy logic as a powerful and potentially useful tool.

Basically, the rationale behind Zadeh's (1973) linguistic approach to the analysis of complex systems and decision (and control) processes is that the basic element is a *linguistic variable* exemplified by “temperature” which takes on as their values not conventional numerical values as, e.g., 150 °C, but linguistic values as, e.g., “high,” “low,” etc. that are in turn equated semantically with some fuzzy sets. Notice that such linguistic values are common in human discourse as natural language is the only fully natural human means of communication. Clearly, one can then form more complex linguistic expressions as, e.g., “not very low and not very high,” “more or less medium,” etc. by using some connectives (e.g., and, or, ...), modifiers (e.g., more or less, very, ...), etc. (see also section “[Fuzzy Logic: Basic Issues](#)”), and employing a syntactic analysis.

To represent a relationship between linguistic variables, *fuzzy conditional statements* are employed. For instance, if we have two linguistic variables, a primary one L and a secondary one K , such that the value of L is a fuzzy set A in X , and the value of K is a fuzzy set B in Y , then a relationship between L and K , in terms of their values A and B , respectively, may be written as

$$\text{IF } L = A \text{ THEN } K = B \quad (52)$$

or, shortly

$$\text{IF } A \text{ THEN } B. \quad (53)$$

This fuzzy conditional statement is now assumed to be equivalent to

$$\text{IF } A \text{ THEN } B = A \times B, \quad (54)$$

i.e., to the Cartesian product (49) of the two fuzzy sets A and B which is in turn a fuzzy relation in $X \times Y$.

Notice that this is what Mamdani (1974) used in his controller, and which is often called Mamdani's implication (though it is not an implication!). This definition is the simplest one, and we can devise more sophisticated ones by using, e.g., various definitions of implication (80, 81, 82, 83, 84, and 85).

It is easy to see that the fuzzy conditional statement (53) may be extended to account for multiple values of A and B obtaining, if we use (54),

$$\begin{aligned} &\text{IF } A_1 \text{ THEN } B_1 \text{ ELSE } \dots \text{ ELSE IF } A_n \text{ THEN } B_n \\ &= A_1 \times B_1 + \dots + A_n \times B_n, \end{aligned} \quad (55)$$

where A_i 's are fuzzy sets in X and B_i 's are fuzzy sets in Y , $i = 1, \dots, n$.

Notice that in (53) and (55) we only specify what happens if the primary variable takes on some value. It is often, however, also relevant to explore what happens if that value is not taken. In such a case, (53) becomes

$$\text{IF } A \text{ THEN } B \text{ ELSE } C, \quad (56)$$

and it is represented as

$$\begin{aligned} &\text{IF } A \text{ THEN } B \text{ ELSE } C \\ &= A \times B + \neg A \times C. \end{aligned} \quad (57)$$

Evidently, one can generalize the above fuzzy conditional statements to involve more than one primary variable. For details, we refer the reader to, e.g., Zadeh (1973).

We therefore have some tool to represent a relation between a primary and secondary variable that is represented by some fuzzy relation. An immediate question is then:

If the primary variable takes on some fuzzy value, say A' , and we have a (fuzzy) relation, IF A THEN B , then what will be the implied (inferred) value of the secondary variable B' ?

This may be represented by the inference scheme

$$\begin{aligned} &A' \\ &\underline{\text{IF } A \text{ THEN } B} \\ &B' = ? \end{aligned} \quad (58)$$

and, what is the very essence, the fuzzy values A' and A need not be the same (notice that this prohibits the use of conventional logical inference tools).

The answer to the question (58) is provided by the *compositional rule of inference* which states that if R in $X \times Y$ is a fuzzy relation representing a dependence between a primary and secondary variable, represented by a fuzzy conditional statement, and the primary variable takes on a fuzzy value A' in X , then the implied fuzzy value of the secondary variable B' in Y is given by the (max-min) composition (47) of A' and R , i.e.,

$$\mu_{B'}(y) = \max_{x \in X} [\mu_{A'}(x) \wedge \mu_R(x, y)], \text{ for each } y \in Y, \quad (59)$$

and notice that A' is here considered to be a unary fuzzy relation as mentioned in section "Fuzzy Relations."

Example 15 Suppose that $X = \{x\} = \{1, 2, 3\}$ and $Y = \{y\} = \{1, 2, 3, 4\}$, and the fuzzy conditional statement representing the dependence between L and K is

$$\text{IF } L = \text{low THEN } K = \text{high} \\ = (\text{low}) \circ (\text{high})$$

where

$$\text{low} = 1/1 + 0.7/2 + 0.3/3 \\ \text{high} = 0.2/1 + 0.5/2 + 0.8/3 + 1/4$$

and is equivalent to the following fuzzy relation

$$R = (\text{"low"}) \circ (\text{"high"}) =$$

	$y = 1$	2	3	4
$x = 1$	0.2	0.5	0.8	1
2	0.2	0.5	0.7	0.7
3	0.2	0.3	0.3	0.3

If now $L = \text{medium} = 0.5/1 + 1/2 + 0.5/3$, then the K induced is given by

$$K = (\text{medium}) \circ R = \max_{x \in \{1, 2, 3\}} [\mu_L(x) \wedge \mu_R(x, y)] = 0.2/1 + 0.5/2 + 0.7/3 + 0.7/4.$$

The fuzzy conditional statements may be used to represent simple dependencies and relations between linguistic variables. For more complicated dependencies and relations, fuzzy algorithms may be used (cf. Zadeh (1973)) which will not be dealt with here.

A further extension in the spirit of the linguistic approach presented in this section is Zadeh's *computing with words* or, more generally, *computing with words and perceptions*. We will not deal with this and will refer the reader to a comprehensive coverage of main theoretical issues, and many applications, related to this approach which is given in Zadeh and Kacprzyk (1999a, b).

The Extension Principle

Now we will briefly present the essence of Zadeh's classic *extension principle* (cf. Zadeh

(1973)) which is one of the most important and powerful tools in fuzzy sets theory. The extension principle addresses the following fundamental issue:

If there is some relationship (e.g., a function) between *conventional* (nonfuzzy) entities (e.g., variables taking on nonfuzzy values), then what is its equivalent relationship between fuzzy entities (e.g., variables taking on fuzzy values)?

The extension principle makes it therefore possible, for instance, to extend some known conventional models, algorithms, etc. involving nonfuzzy variables to the case of fuzzy variables.

Let $A_1, \dots, A_n$ be fuzzy sets in $X_1 = \{x_1\}, \dots, X_n = \{x_n\}$, respectively, and

$$f : X_1 \times \dots \times X_n \rightarrow Y \quad (60)$$

be some (nonfuzzy) function such that $y = f(x_1, \dots, x_n)$.

Then, according to the *extension principle*, the fuzzy set B in $Y = \{y\}$ induced by the fuzzy sets $A_1, \dots, A_n$ via the function f (60) is

$$\mu_B(y) = \max_{(x_1, \dots, x_n) \in X_1 \times \dots \times X_n : y = f(x_1, \dots, x_n)} \bigwedge_{i=1}^n \mu_{A_i}(x_i). \quad (61)$$

Example 16 Suppose that: $X_1 = \{1, 2, 3\}$, $X_2 = \{1, 2, 3, 4\}$, f represents the addition, i.e., $y = x_1 + x_2$, $A_1 = 0.1/1 + 0.6/2 + 1/3$, and $A_2 = 0.6/1 + 1/2 + 0.5/3 + 0.1/4$, then

$$B = A_1 + A_2 \\ = 0.1/2 + 0.6/3 + 0.6/4 + 1/5 + 0.5/6 \\ + 0.1/7$$

and notice that $+$ is used here in both the arithmetic (the sum of real and fuzzy numbers – cf. section “Fuzzy Numbers”) and set-theoretic sense which should not lead to confusion.

Equivalently, the extension principle (61) may also be written in terms of the α -cuts (18). Namely, suppose for simplicity that $f : X \rightarrow Y$, $X = \{x\}$, $Y = \{y\}$, and A_α , for each $\alpha \in (0, 1]$, are α -cuts of

A . Then, the fuzzy set B in Y , induced by A via the extension principle, is given as

$$\begin{aligned} B = f(A) &= f\left(\sum_{\alpha \in (0,1]} \alpha \cdot A_\alpha\right) \\ &= \sum_{\alpha} \alpha f(A_\alpha) \end{aligned} \quad (62)$$

which is clearly implied by the representation theorem (Theorem 1).

Notice that the extension principle plays an extraordinary role in the sense that we have a multitude of nonfuzzy tools like algorithms, procedures, etc. which have been widely used to solve various problems. However, they need precise information, notably real or integer numbers, real intervals, etc. We have neither fuzzy computers nor fuzzy tools of the type mentioned above so that we are not in a position to directly use imprecise information to solve our problems. However, by virtue of the extension principle we can extend our known nonfuzzy tools and techniques (their related algorithms and procedures) to their fuzzy counterparts.

Fuzzy Numbers

The same fundamental role as played by nonfuzzy (real, integer, ...) numbers in conventional models, the fuzzy numbers play in fuzzy models.

A *fuzzy number* is defined as a fuzzy set in R , the real line. Usually, but not always, it is assumed to be a normal and convex fuzzy set. For example, the membership function of a fuzzy number that is *more or less six* may be as shown in Fig. 6, i.e., as a bell-shaped function.

For our purposes, operations on fuzzy numbers are the most relevant. It is easy to notice that their definitions may readily be obtained by applying the extension principle (61).

Suppose therefore that A and B are two fuzzy numbers in $R = \{x\}$ characterized by their membership functions $\mu_A(x)$ and $\mu_B(x)$, respectively. Then the extension principle yields the following definitions of the four basic *arithmetic operations* on fuzzy numbers:

- *Addition*

$$\begin{aligned} \mu_{A+B}(z) &= \max_{x+y=z} [\mu_A(x) \wedge \mu_B(y)], \\ &\text{for each } z \in R, \end{aligned} \quad (63)$$

- *Subtraction*

$$\begin{aligned} \mu_{A-B}(z) &= \max_{x-y=z} [\mu_A(x) \wedge \mu_B(y)], \\ &\text{for each } z \in R \end{aligned} \quad (64)$$

- *Multiplication*

$$\begin{aligned} \mu_{A \cdot B}(z) &= \max_{x \cdot y=z} [\mu_A(x) \wedge \mu_B(y)], \\ &\text{for each } z \in R \end{aligned} \quad (65)$$

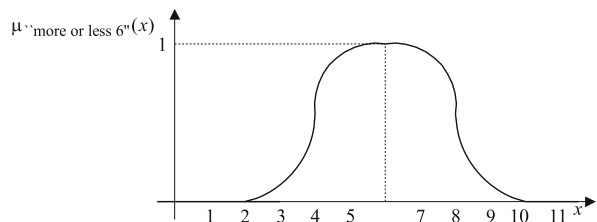
- *Division*

$$\begin{aligned} \mu_{A/B}(z) &= \max_{x/y=z, y \neq 0} [\mu_A(x) \wedge \mu_B(y)], \\ &\text{for each } z \in R \end{aligned} \quad (66)$$

In some applications, the following one-argument operations on fuzzy numbers may also be of use:

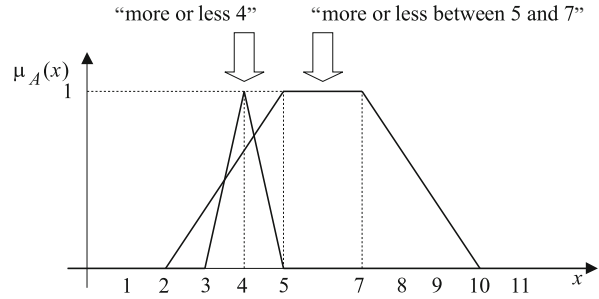
- The *opposite* of a fuzzy number

Foundations of Fuzzy Sets Theory, Fig. 6 The membership function of a fuzzy number *more or less six*



Foundations of Fuzzy Sets Theory,

Fig. 7 Triangular (*about four*) and trapezoid (*more or less between six and seven*) fuzzy numbers



$$\mu_{-A}(x) = \mu_A(-x), \quad \text{for each } x \in R \quad (67)$$

- The *inverse* of a fuzzy number

$$\mu_{A^{-1}}(x) = \mu_A\left(\frac{1}{x}\right), \quad \text{for each } x \in R \setminus \{0\}. \quad (68)$$

In practice, however, such a general definition of fuzzy numbers and operations on them is seldom used. Normally a further simplification is made, namely the fuzzy numbers are assumed to be *triangular* and eventually *trapezoid* fuzzy numbers whose membership functions are sketched in Fig. 7.

For triangular and trapezoid fuzzy numbers, for the four basic operations, i.e., the addition, subtraction, multiplication, and division, special formulas can be devised whose calculation is simpler than that of (63–66). Moreover, a crucial problem of comparison of two fuzzy numbers may also be simplified. For details, we will refer the reader to a rich literature exemplified by Dubois and Prade (1980), Kaufmann and Gupta (1985), Klir and Folger (1988), Klir and Yuan (1995), Maresš (1994), Hanss (2005), etc.

Fuzzy Events and Their Probabilities

Fuzziness and randomness are meant, in our perspective, as two aspects of *imperfect information*. Fuzziness is meant to concern entities and relations which are not crisply defined, with gradual transition between the elements belonging and not belonging to a class. Randomness concerns situations in which the event is well defined, but its occurrence is uncertain.

However, in practice and in everyday discourse there is an abundance of situations in

which there jointly occur fuzziness and randomness, for instance, when we ask about the probability of *cold weather* tomorrow or of a *high inflation* in the next year, we have an imprecise (fuzzy) event – cold weather and high inflation, respectively. To be able to formally deal with such problems, we need a concept of a fuzzy event, and that of a probability of a fuzzy event.

The first approach in this respect is due to Zadeh (1968). Its point of departure is the concept of a *fuzzy event* which is simply a fuzzy set A in $X = \{x\} = \{x_1, \dots, x_n\}$ whose membership function is Borel measurable. We assume that the probabilities of the (nonfuzzy) elementary events $x_1, \dots, x_n \in X$ are known and equal to $p(x_1), \dots, p(x_n) \in [0, 1]$, respectively, with $p(x_1) + \dots + p(x_n) = 1$.

As to some more important concepts related to fuzzy events, the following may be stated.

Two fuzzy events A and B in X are *independent* if and only if

$$p(AB) = p(A)p(B) \quad (69)$$

The *conditional probability* of a fuzzy event A in X with respect to a fuzzy event B in X is denoted $p(A | B)$ and defined as

$$p(A | B) = \frac{p(AB)}{p(B)}, \quad p(B) > 0, \quad (70)$$

and if the fuzzy events A and A are independent, then

$$p(A | B) = p(A). \quad (71)$$

Notice that both of the above concepts are analogous to their nonfuzzy counterparts.

The (nonfuzzy) probability of a fuzzy event A in $X = \{x_1, \dots, x_n\}$ is denoted $p(A)$ and defined by Zadeh (1968) as

$$p(A) = \sum_{i=1}^n \mu_A(x_i) p(x_i), \quad (72)$$

i.e., as the expected value of the membership function of A , $\mu_A(x)$.

Example 17 Suppose that $X = \{1, 2, \dots, 5\}$, $p(x_1) = 0.1$, $p(x_2) = 0.1$, $p(x_3) = 0.1$, $p(x_4) = 0.3$, $p(x_5) = 0.4$, and $A = 0.1/2 + 0.5/3 + 0.7/4 + 0.9/5$.

Then

$$p(A) = 0.1 \times 0.1 + 0.1 \times 0.5 + 0.3 \times 0.7 + 0.4 \times 0.9 = 0.73.$$

Notice that Zadeh's (1968) (nonfuzzy) probability of a fuzzy event (72) satisfies the following:

1. $p(\emptyset) = 0$
2. $p(\neg A) = 1 - p(A)$
3. $p(A + B) = p(A) + p(B) - p(A \cap B)$
4.
$$p\left(\sum_{i=1}^r A_i\right) = \sum_{i=1}^r p(A_i) - \sum_{j=1}^r \sum_{k=1, k < j}^r p(A_j \cap A_k) \\ + \sum_{j=1}^r \sum_{k=1, k < j}^r \sum_{l=1, l < k}^r p(A_j \cap A_k \cap A_l) \\ + \dots + (-1)^{r+1} p(A_1 \cap A_2 \cap \dots \cap A_r)$$

So it does make sense to term the expression (72) a "probability."

The above Zadeh's (1968) classic definition of a (nonfuzzy) probability of a fuzzy event is by far the most popular and most widely used. However, though the event is fuzzy, its probability is nonfuzzy, i.e., is a real number from the unit interval. This may be viewed counterintuitive but provides simplicity. For some approaches to a fuzzy probability of a fuzzy event, see Klir and Folger (1988) or Klir and Yuan (1995).

Defuzzification of Fuzzy Sets

In many applications, we arrive at a fuzzy result. However, in it is a crisp (nonfuzzy) result that should be applied. A notable example is fuzzy control (cf. Driankov, Hellendoorn, and Reinfrank (1993) or Kacprzyk (1996)).

Suppose that we have a fuzzy set A defined in $X = \{x_1, x_2, \dots, x_n\}$, i. e. $A = \mu_A(x_1)/x_1 + \mu_A(x_2)/x_2 + \dots + \mu_A(x_n)/x_n$. We need to find a crisp number $a \in [x_1, x_n]$ which best represents A . Notice that we assume here that A is defined in a finite universe of discourse, but its corresponding defuzzified number a need not be in general any of the finite values of X but should be between the lowest and highest elements of X (evidently, this requires some ordering of x_i 's, but this is clearly satisfied as x_i 's 0.33 are in virtually all practical cases just real numbers).

The most commonly used defuzzification procedure is certainly the *center-of-area*, also called the *center-of-gravity*, method whose essence is

$$a = \frac{\sum_{i=1}^n x_i \mu_A(x_i)}{\sum_{i=1}^n \mu_A(x_i)} \quad (73)$$

The above defuzzification (73) is however often too complex if our analysis involves, e.g., some optimization (cf. Kacprzyk (1996)). In such a case, one needs to resort to an even simpler defuzzification method which simply assumes that the defuzzified value of a fuzzy value is $x_i \in X = \{x_1, \dots, x_n\}$ for which $\mu_A(x)$ takes on its maximum values, i.e.,

$$\mu_A(a) = \max_{x_i \in X} \mu_A(x) \quad (74)$$

with an obvious extension that if the A determined in (74) is not unique, then we take, say, the mean value of such equivalent a 's.

In section "Bellman and Zadeh's General Approach to Decision Making Under Fuzziness," we will provide a justification of the above maximizing-value-type defuzzification procedure in the framework of decision-making.

There is a whole array of other defuzzification procedures, and the reader is referred to, e.g., Driankov, Hellendorn, and Reinfrank (1993), Kacprzyk (1996), Klir and Yuan (1995), or Yager and Filev (1994).

Fuzzy Logic: Basic Issues

The concept of a *fuzzy logic* is not uniquely understood. Basically, it may be meant in (at least) the three following ways:

- As a foundation of reasoning based on ambiguous, vague, and imprecise statements (cf. Goguen (1969))
- As a foundation of reasoning based on ambiguous, vague, and imprecise statements in which fuzzy set-theoretic tools are used (see Zadeh (1975a); or Zadeh and Kacprzyk (1992))
- As a multivalued logic with truth values in the unit interval in which the logical operations of negation, union, intersection, implication, equivalence, etc. are chosen in a special way, and have some fuzzy interpretation (cf. Hájek (1998) or Nová, Perfilieva, and Močkoř (1999))

It is easy to notice that the meaning of fuzzy logic in the first and second sense is similar, though the generality of the former is clearly higher, while its meaning in the third way is different.

For the purposes of this entry, we will assume the third view on fuzzy logic and will present a very limited survey of basic issues. The interested reader is referred for more detail on various aspects of fuzzy logic to Zadeh and Kacprzyk's (1992) volume which is practically the only up-to-date and exhaustive treatise on fuzzy logic available today.

Notice what some authors mean by fuzzy logic is the whole theory of fuzzy sets and related topics. We will not follow this line of reasoning, and view fuzzy sets theory as more *set-theoretic* while fuzzy logic as more *logical*.

Suppose that we have a statement (predicate) u is P denoted, for brevity, as P and exemplified

by temperature (u) is high (P), where u is a variable taking on its values in a universe of discourse $U = \{u\}$, and P is an imprecise term equated with a fuzzy set in U , $P = \{\mu_P(u)/u\}$.

For a specified value $u \in U$, the truth of u is P (or of P) is denoted $\tau(P)$ and meant to be $\tau(u \text{ is } P) = \tau(P) = \mu_P(u)$, for each $u \in U$.

The following general definitions of basic logical operations (in terms of their respective truth values) are usually employed:

- The *negation* of P , i.e., *not* P , denoted by $\neg P$:

$$\tau(\neg P) = 1 - \tau(P). \quad (75)$$

- The *intersection* of P and Q , i.e., P and Q , denoted by $P \cap Q$:

$$\tau(P \cap Q) = t[\tau(P), \tau(Q)]. \quad (76)$$

where $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a t -norm (36); the original Zadeh's definition is

$$\tau(P \cap Q) = \tau(P) \wedge \tau(Q) = \min [\tau(P), \tau(Q)]. \quad (77)$$

- The *union* of P and Q , i.e., P or Q , denoted by $P \cup Q$:

$$\tau(P \cup Q) = s[\tau(P), \tau(Q)] \quad (78)$$

where $[0, 1] \times [0, 1] \rightarrow [0, 1]$ is an s -norm (40); the original Zadeh's definition is

$$\tau(P \cup Q) = \tau(P) \vee \tau(Q) = \max [\tau(P), \tau(Q)]. \quad (79)$$

- The *implication*, i.e., if P then Q , denoted by $P \Rightarrow Q$, which may be defined as, e.g.,

1. The Łukasiewicz implication

$$\tau(P \Rightarrow Q) = \min \{1 - \tau(P) + \tau(Q), 1\} \quad (80)$$

2. The Gödel implication

$$Qy' \text{ s are } F, \quad (87)$$

$$\tau(P \Rightarrow Q) = \begin{cases} 1 & \text{if } \tau(P) \leq \tau(Q) \\ \tau(Q) & \text{otherwise} \end{cases} \quad (81)$$

3. The Goguen implication

$$\tau(P \Rightarrow Q) = \begin{cases} 1 & \text{if } \tau(P) = 0 \\ \min \left\{ 1, \frac{\tau(Q)}{\tau(P)} \right\} & \text{otherwise} \end{cases} \quad (82)$$

4. The Kleene-Dienes implication

$$\tau(P \Rightarrow Q) = \max \{1 - \tau(P), \tau(Q)\} \quad (83)$$

5. The Zadeh implication

$$\tau(P \Rightarrow Q) = \max \{1 - \tau(P), \min \{\tau(P), \tau(Q)\}\} \quad (84)$$

6. The Reichenbach implication

$$\tau(P \Rightarrow Q) = 1 - \tau(P) + \tau(P) \cdot \tau(Q) \quad (85)$$

- The *equivalence*, i.e., P is equivalent to Q or if P then Q and if Q then P , denoted by $P \Leftrightarrow Q$:

$$\tau(P \Leftrightarrow Q) = \tau[(P \Rightarrow Q) \cap (Q \Rightarrow P)] \quad (86)$$

- And we can assume an appropriate definition of the intersection $\cap$ (34), and the implication $\Rightarrow$ (80, 81, 82, 83, 84, and 85).

Among some other important aspects of fuzzy logic, one should mention the use of (fuzzy) linguistic quantifiers, exemplified by *most*, *almost all*, *a few*, etc., which are common in everyday discourse but cannot be handled by conventional logic in which only the two quantifiers are in principle employed, i.e., the universal quantifier *for all* and the existential quantifier *for at least one*.

A *linguistically quantified statement* is exemplified by *most experts are convinced* and may be generally written as

where Q is a linguistic quantifier (e.g., *most*), $Y = \{y\}$ is a set of objects (e.g., experts), and F is a property (e.g., convinced).

Importance B may also be added to the linguistically quantified statement (87) yielding

$$QBy' \text{ s are } F \quad (88)$$

exemplified by *most (Q) of the important (B) experts (y's) are convinced (F)*.

For our purposes, the problem is to find the (degree of) truth of such linguistically quantified statements (87) and (88), denoted $\tau(Qy' \text{ s are } F)$ in the former case and $\tau(QBy' \text{ s are } F)$ in the latter case, knowing the truth of the statements, y is F , denoted $\tau(y \text{ is } F)$, for all $y \in Y$. Evidently, all these degrees of truth (truths) will be meant as real numbers from the unit interval.

Fortunately enough, these truth values may be found by some fuzzy logic calculi [cf. Kacprzyk (1996), Yager (1983), or Zadeh (1983)]. For lack of space, we are unable to discuss these issues here and refer the reader to, e.g., the source papers or Kacprzyk's (1996) book.

Fuzzy linguistic quantifiers are very relevant both for theory and applications (ranging from decision analysis, social choice, optimization, and control to database queries). The fuzzy linguistic quantifiers have a relation to the so-called ordered weighted averaging (OWA) operators introduced by Yager in 1982, and we refer the reader for a comprehensive coverage of their theory and applications to Yager and Kacprzyk's (1996) volume.

Bellman and Zadeh's General Approach to Decision-Making Under Fuzziness

Fuzzy logic, the essence of which has been presented in previous sections, has found applications in a multitude of areas of science and technology, and a full coverage is beyond the scope of our exposition.

Since *decision-making* is by far the most well-known, omnipresent problem, we will just sketch the application of fuzzy sets theory to the broadly perceived decision-making.

The purpose of this section is to provide the reader with a brief introduction to Bellman and Zadeh's (1970) general approach to decision-making under fuzziness, originally termed *decision-making in a fuzzy environment*, a simple yet extremely powerful framework within which virtually all fuzzy models related to decision-making, optimization, and control have been dealt with.

In Bellman and Zadeh's (1970) setting, the imprecision (fuzziness) of the environment within which the decision-making (control, ...) process proceeds is modeled by the introduction of the so-called *fuzzy environment* which consists of fuzzy goals, fuzzy constraints, and fuzzy decision.

Suppose that we have some set of possible options (or alternatives, variants, choices, decisions, ...), $X = \{x\}$, which contains all the possible (relevant, feasible, ...) values, courses of action, etc.

The *fuzzy goal* is now defined as a fuzzy set in the set G in the set of options X , characterized by its membership function $\mu_G: X \rightarrow [0, 1]$ such that $\mu_G(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in the fuzzy goal G .

The *fuzzy constraint* is similarly defined as a fuzzy set C in the set of options X , characterized by its membership function $\mu_C: X \rightarrow [0, 1]$ such that $\mu_C(x) \in [0, 1]$ specifies the grade of membership of a particular option $x \in X$ in the fuzzy constraint C .

The fuzzy goal and fuzzy constraint are illustrated in Example 18.

As to the interpretation of fuzzy goals and/or constraints (notice that their definitions do indicate an intrinsic analogy!), in some, mostly earlier, works as, e.g., in Bellman and Zadeh (1970), the following view on the essence of the fuzzy goal is advocated. Suppose that $f: X \rightarrow R$ is a conventional performance (objective) function which associates with each option $x \in X$ a real number $f(x) \in R$, and which is bounded, i.e., $f(x) \leq M < \infty$, for each $x \in X$, where $M = \max_{x \in X} f(x)$.

Then the membership function of the fuzzy goal G can be defined as a normalized performance function f , i.e.,

$$\begin{aligned}\mu_G(x) &= \frac{f(x)}{M} \\ &= \frac{f(x)}{\max_{x \in X} f(x)}, \text{ for each } x \in X. \quad (89)\end{aligned}$$

A fuzzy goal may however be also viewed from a different perspective that is often convenient, i.e., in terms of Simon's *satisfaction levels*, in particular in view of the representation of the fuzzy goal's membership function in a piecewise linear form as usually assumed, also here. Then the piecewise linear membership function of a fuzzy goal in Fig. 8 should be understood as follows: If the value of x attained is at least $\bar{x}_G (= 8)$, which is the *satisfaction level* of x , i.e. for $x \geq \bar{x}_G$, then $\mu_G(x) = 1$ which means that we are fully satisfied with the x attained. On the other hand, if the x attained does not exceed $\underline{x}_G (= 5)$, which is the lowest possible value of x , then $\mu_G(x) = 0$ which means that we are fully dissatisfied with such a value of x or, in other words, this value is impossible. For the intermediate values, $\underline{x}_G < x < \bar{x}_G$, we have $0 < \mu_G(x) < 1$ which means that our satisfaction as to a particular value of x is intermediate. The interpretation of a fuzzy constraint is analogous.

It is now easy to see that the above interpretation provides a *common denominator* for the fuzzy goal and fuzzy constraint. They may be treated in an analogous way which is one of the merits of Bellman and Zadeh's (1970) approach.

The above suggests the following general formulation of the decision-making problem in a fuzzy environment

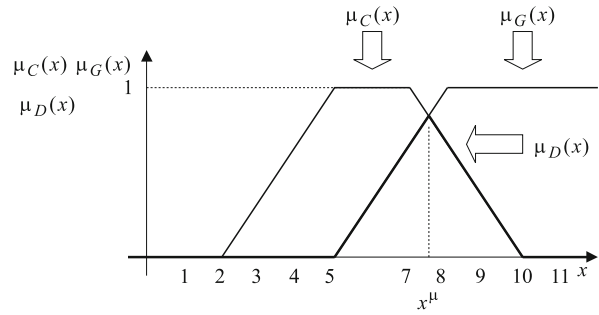
$$\text{"Attain } G \text{ and satisfy } C" \quad (90)$$

which should be meant as to determine a decision (an option or a set of options) which simultaneously fulfills the fuzzy goal and fuzzy constraint, and which belongs to the available (or, maybe, relevant, feasible, ...) ones.

The fuzziness of the fuzzy goal and fuzzy constraint implies the above decision, a fuzzy

Foundations of Fuzzy Sets Theory,

Fig. 8 Fuzzy goal, fuzzy constraint, fuzzy decision, and the optimal (maximizing) decision



decision, to be a fuzzy set defined in the set of options which results from the intersection (34) of the fuzzy goal and fuzzy constraint. Formally, if G is a fuzzy goal and C is a fuzzy constraint, both defined as fuzzy sets in the set of options $X = \{x\}$, the *fuzzy decision* D is a fuzzy set defined in X given as

$$\mu_D(x) = \mu_G(x) \wedge \mu_C(x), \text{ for each } x \in X \quad (91)$$

where $\wedge$ is the minimum operation, i.e., $a \wedge b = \min(a, b)$.

The fuzzy decision (91) is most widely used, but $\wedge$ may clearly be replaced by another operation as, e.g., a t -norm (36).

Example 18 Suppose that G is *x should be much larger than five*, and C is *x should be about 5*, as in Fig. 8.

The membership function of the fuzzy decision is given in heavy line and interpreted as follows. The set of possible options is the interval (Bandemer and Näther 1992; Bezdek 1981) because $\mu_D(x) > 0$ for $5 \leq x \leq 10$. The value of $\mu_D(x) \in [0, 1]$ is meant as the degree of satisfaction from the choice of a particular $x \in X$, from zero for full dissatisfaction (impossibility of x) to one for full satisfaction, though all intermediate values, and the higher the value of $\mu_D(x)$, the higher the satisfaction from x .

Notice that in Fig. 8, $\mu_D(x) < 1$ which means that there is no option which fully satisfies both the fuzzy goal and fuzzy constraint. In other words, there is a discrepancy or conflict between the fuzzy goal and constraint.

In practice, however, we need to find a non-fuzzy solution to be implemented. The above interpretation of the fuzzy decision immediately suggests that the best (nonfuzzy) choice in this case would be the one corresponding to the highest value of $\mu_D(x)$.

The *maximizing decision* is therefore defined as an $x^* \in X$ such that

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x) \quad (92)$$

and an example may be found in Fig. 8 where $x^* = 7.5$.

Notice that the above is clearly equivalent to the defuzzification of the fuzzy decision (cf. section “Defuzzification of Fuzzy Sets”), and other defuzzification procedures may also be used in principle. However, (92) is often the only practical choice (cf. Kacprzyk (1996)) and will be assumed here.

In nontrivial real problems, there are multiple fuzzy goals and fuzzy constraints, and they may be handled within the above framework in quite a straightforward manner.

Suppose a more general situation than the fuzzy constraint C is defined as a fuzzy set in $X = \{x\}$, and the fuzzy goal G is defined as a fuzzy set in $Y = \{y\}$. Moreover, suppose that a function $f: x \rightarrow Y, y = f(x)$ is known. Typically, X and Y may be sets of options and outcomes, notably causes and effects.

Now, the *induced fuzzy goal* G' in X generated by the given fuzzy goal G in Y is defined as

$$\mu_{G'}(x) = \mu_G[f(x)], \text{ for each } x \in X. \quad (93)$$

Example 19 Let $X = \{1, 2, 3, 4\}$, $Y = \{2, 3, \dots, 10\}$, and $y = 2x + 1$. If now

$$G = 0.1/2 + 0.2/3 + 0.4/4 + 0.5/5 + 0.6/6 \\ + 0.7/7 + 0.8/8 + 1/9 + 1/10$$

then

$$G' = \mu_G(3)/1 + \mu_G(5)/2 + \mu_G(7)/70.2/1 + 0.5/2 \\ + 0.7/3 + 1/4$$

The *fuzzy decision* is now defined analogously as (91), i.e.,

$$\mu_D(x) = \mu_{G'}(x) \wedge \mu_C(x), \text{ for each } x \in X. \quad (94)$$

Clearly, for $n > 1$ fuzzy goals $G_1, \dots, G_n$ defined in Y , $m > 1$ fuzzy constraints $C_1, \dots, C_m$ defined in X , and a function $f: X \rightarrow Y, y = f(x)$, we analogously have

$$\mu_D(x) = \mu_{G'_1}(x) \wedge \dots \wedge \mu_{G'_n}(x) \wedge \mu_{C_1}(x) \\ \wedge \dots \wedge \mu_{C_m}(x), \text{ for each } x \in X. \quad (95)$$

The *maximizing decision* is defined as (92), i.e.,

$$\mu_D(x^*) = \max_{x \in X} \mu_D(x).$$

The basic conceptual fuzzy decision-making model can be used in many specific areas, notably

in fuzzy optimization which will be covered elsewhere in this volume.

The models of decision-making under fuzziness developed above can also be extended to the case of multiple criteria, multiple decision makers, and multiple stage cases. We will present the last extension to fuzzy multistage decision-making (control) case which makes it possible to account for dynamics.

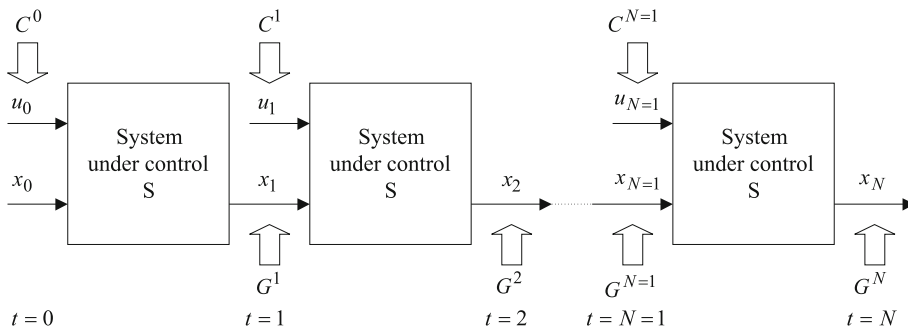
Multistage Decision-Making (Control) Under Fuzziness

In this case, it is convenient to use control-related notation and terminology. In particular, decisions will be referred to as *controls*, the discrete time moments at which decisions are to be made – as *control stages*, and the input-output (or cause-effect) relationship – as a *system under control*.

The essence of multistage control in a fuzzy environment may be portrayed as in Fig. 9.

First, suppose that the control space is $U = \{u\} = \{c_1, \dots, c_m\}$ and the state space is $X = \{x\} = \{s_1, \dots, s_n\}$. Initially, we are in some initial state $x_0 \in X$. We apply a control $u_0 \in U$ subjected to a fuzzy constraint $\mu_{C^0}(u_0)$. We attain a state $x_1 \in X$ via a known cause-effect relationship (i.e., S); a fuzzy goal $\mu_{G^1}(x_1)$ is imposed on x_1 . Next, we apply a control u_1 subjected to a fuzzy constraint $\mu_{C^1}(u_1)$ and attain a fuzzy state x_2 on which a fuzzy goal $\mu_{G^2}(x_2)$ is imposed, etc.

Suppose for simplicity that the system under control is deterministic and its temporal evolution is governed by a *state transition equation*



Foundations of Fuzzy Sets Theory, Fig. 9 Essence of the multistage control in a fuzzy environment (under fuzziness)

$$f : X \times U \rightarrow X, \quad (96)$$

such that

$$x_{t+1} = f(x_t, u_t), t = 0, 1, \dots \quad (97)$$

where $x_t, x_{t+1} \in X = \{s_1, \dots, s_n\}$ are the states at control stages t and $t + 1$, respectively, and $u_t \in U = \{c_1, \dots, c_m\}$ is the control at t .

At each t , $u_t \in U$ is subjected to a fuzzy constraint $\mu_{C^t}(u_t)$, and on the state attained $x_{t+1} \in X$ a fuzzy goal is imposed; $t = 0, 1, \dots$. The initial state is $x_0 \in X$ and is assumed to be known, and given in advance. The termination time (planning, or control, horizon), i.e., is the maximum number of control stages, is denoted by $N \in \{1, 2, \dots\}$ and may be finite or infinite.

The performance (goodness) of the multistage control process under fuzziness is evaluated by the fuzzy decision

$$\begin{aligned} \mu_D(u_0, \dots, u_{N-1}|x_0) &= \mu_{C^0}(u_0) \wedge \mu_{G^1}(x_1) \wedge \dots \\ &\wedge \mu_{C^{N-1}}(u_{N-1}) \wedge \mu_{G^N}(x_N). \end{aligned} \quad (98)$$

In most cases, however, a slightly simplified form of the fuzzy decision (98) is used, namely it is assumed all the subsequent fuzzy controls, $u_0, u_1, \dots, u_{N-1}$, are subjected to the fuzzy constraints, $\mu_{C^0}(u_0), \mu_{C^1}(u_1), \dots, \mu_{C^{N-1}}(u_{N-1})$, while the fuzzy goal is just imposed on the final state x_N , $\mu_{G^N}(x_N)$. In such a case, the fuzzy decision becomes

$$\begin{aligned} \mu_D(u_0, \dots, u_{N-1}|x_0) &= \mu_{C^0}(u_0) \wedge \dots \\ &\wedge \mu_{C^{N-1}}(u_{N-1}) \wedge \mu_{G^N}(x_N). \end{aligned} \quad (99)$$

The multistage control problem in a fuzzy environment is now formulated as to find an optimal sequence of controls $u_0^*, \dots, u_{N-1}^*, u_t^* \in U, t = 0, 1, \dots, N - 1$, such that:

$$\begin{aligned} &\mu_D(u_0^*, \dots, u_{N-1}^*|x_0) \\ &= \max_{u_0, \dots, u_{N-1} \in U} \mu_D(u_0, \dots, u_{N-1}|x_0). \end{aligned} \quad (100)$$

Usually it is more convenient to express the solution, i.e., the controls to be applied, as a control policy $a_t : X \rightarrow U$, such that $u_t = a_t(x_t), t = 0, 1, \dots$, i.e., the control to be applied at t is expressed as a function of the state at t .

The above basic formulation of multistage control in a fuzzy environment may readily be extended with respect to:

- The type of the termination time (fixed and specified, implicitly specified, fuzzy, and infinite)
- The type of the system under control (deterministic, stochastic, and fuzzy)

For a detailed analysis of resulting problems and their solutions (by employing dynamic programming, branch-and-bound, genetic algorithms, etc.), we refer the reader to Kacprzyk's (1983, 1996) books.

Concluding Remarks

We provided a brief survey of basic elements of Zadeh's (1965) fuzzy sets theory, mainly of basic properties of fuzzy sets, operations on fuzzy sets, fuzzy relations and their compositions, linguistic variables, the extension principle, fuzzy arithmetic, fuzzy events and their probabilities, fuzzy logic, and Bellman and Zadeh's (1970) general approach to decision-making in a fuzzy environment. Various aspects of fuzzy sets theory will be expanded in other papers in this part of the volume.

Bibliography

- Atanassov KT (1983) Intuitionistic fuzzy sets. VII ITRK session. Central Sci.-Techn. Library of Bulg. Acad. of Sci., Sofia pp 1697/84. (in Bulgarian)
- Atanassov KT (1986) Intuitionistic fuzzy sets. Fuzzy Sets Syst 20:87-96
- Atanassov KT (1999) Intuitionistic fuzzy sets: theory and applications. Springer, Heidelberg
- Bandemer H, Gottwald S (1995) Fuzzy sets, fuzzy logic, fuzzy methods, with applications. Wiley, Chichester
- Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer, Dordrecht

- Bandler W, Kohout LJ (1980) Fuzzy power sets for fuzzy implication operators. *Fuzzy Sets Syst* 4:13–30
- Bellman RE, Giertz M (1973) On the analytic formalism of the theory of fuzzy sets. *Inf Sci* 5:149–157
- Bellman RE, Zadeh LA (1970) Decision making in a fuzzy environment. *Manag Sci* 17:141–164
- Belohlávek R, Vychodil V (2005) *Fuzzy equational logic*. Springer, Heidelberg
- Bezdek JC (1981) *Pattern recognition with fuzzy objective function algorithms*. Plenum, New York
- Black M (1937) Vagueness: an exercise in logical analysis. *Philos Sci* 4:427–455
- Black M (1963) Reasoning with loose concepts. *Dialogue* 2:1–12
- Black M (1970) *Margins of precision*. Cornell University Press, Ithaca
- Buckley JJ (2004) *Fuzzy statistics*. Springer, Heidelberg
- Buckley JJ (2005a) *Fuzzy probabilities. New approach and applications*, 2nd edn. Springer, Heidelberg
- Buckley JJ (2005b) *Simulating fuzzy systems*. Springer, Heidelberg
- Buckley JJ (2006) *Fuzzy probability and statistics*. Springer, Heidelberg
- Buckley JJ, Eslami E (2002) *An introduction to fuzzy logic and fuzzy sets*. Springer, Heidelberg
- Calvo T, Mayor G, Mesiar R (2002) *Aggregation operators. New trends and applications*. Springer, Heidelberg
- Carlsson C, Fullér R (2002) *Fuzzy reasoning in decision making and optimization*. Springer, Heidelberg
- Castillo O, Melin P (2008) *Type-2 fuzzy logic: theory and applications*. Springer, Heidelberg
- Cox E (1994) *The fuzzy system handbook. A practitioner's guide to building, using, and maintaining fuzzy systems*. Academic, New York
- Cross V, Sudkamp T (2002) *Similarity and compatibility in fuzzy set theory. Assessment and applications*. Springer, Heidelberg
- Delgado M, Kacprzyk J, Verdegay JL, Vila MA (eds) (1994) *Fuzzy optimization: recent advances*. Physica, Heidelberg
- Dompere KK (2004a) Cost-benefit analysis and the theory of fuzzy decisions. *Fuzzy value theory*. Springer, Heidelberg
- Dompere KK (2004b) Cost-benefit analysis and the theory of fuzzy decisions. *Identification and measurement theory*. Springer, Heidelberg
- Driankov D, Hellendorn H, Reinfrank M (1993) *An introduction to fuzzy control*. Springer, Berlin
- Dubois D, Prade H (1980) *Fuzzy sets and systems: theory and applications*. Academic, New York
- Dubois D, Prade H (1985) *Théorie des possibilités. Applications à la représentation des connaissances en informatique*. Masson, Paris
- Dubois D, Prade H (1988) *Possibility theory: an approach to computerized processing of uncertainty*. Plenum, New York
- Dubois D, Prade H (1996) *Fuzzy sets and systems (Reedition on CR-ROM of [28])*. Academic, New York
- Fullér R (2000) *Introduction to neuro-fuzzy systems*. Springer, Heidelberg
- Gaines BR (1977) *Foundations of fuzzy reasoning*. Int J Man-Mach Stud 8:623–668
- Gil Aluja J (2004) *Fuzzy sets in the management of uncertainty*. Springer, Heidelberg
- Gil-Lafuente AM (2005) *Fuzzy logic in financial analysis*. Springer, Heidelberg
- Glöckner I (2006) *Fuzzy quantifiers. A computational theory*. Springer, Heidelberg
- Goguen JA (1967) *L-fuzzy sets*. J Math Anal Appl 18: 145–174
- Goguen JA (1969) The logic of inexact concepts. *Synthese* 19:325–373
- Goodman IR, Nguyen HT (1985) *Uncertainty models for knowledge-based systems*. North-Holland, Amsterdam
- Hájek P (1998) *Metamathematics of fuzzy logic*. Kluwer, Dordrecht
- Hanss M (2005) *Applied fuzzy arithmetic. An introduction with engineering applications*. Springer, Heidelberg
- Kacprzyk J (1983) *Multistage decision making under fuzziness*. Verlag TÜV Rheinland, Cologne
- Kacprzyk J (1992) *Fuzzy sets and fuzzy logic*. In: Shapiro SC (ed) *Encyclopedia of artificial intelligence*, vol 1. Wiley, New York, pp 537–542
- Kacprzyk J (1996) *Multistage fuzzy control*. Wiley, Chichester
- Kacprzyk J, Fedrizzi M (eds) (1988) *Combining fuzzy imprecision with probabilistic uncertainty in decision making*. Springer, Berlin
- Kacprzyk J, Orlovski SA (eds) (1987) *Optimization models using fuzzy sets and possibility theory*. Reidel, Dordrecht
- Kandel A (1986) *Fuzzy mathematical techniques with applications*. Addison Wesley, Reading
- Kaufmann A, Gupta MM (1985) *Introduction to fuzzy mathematics – theory and applications*. Van Nostrand Reinhold, New York
- Klement EP, Mesiar R, Pap E (2000) *Triangular norms*. Springer, Heidelberg
- Klir GJ (1987) Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy Sets Syst* 24:141–160
- Klir GJ, Folger TA (1988) *Fuzzy sets, uncertainty and information*. Prentice-Hall, Englewood Cliffs
- Klir GJ, Wierman M (1999) *Uncertainty-based information. Elements of generalized information theory*, 2nd edn. Springer, Heidelberg
- Klir GJ, Yuan B (1995) *Fuzzy sets and fuzzy logic: theory and application*. Prentice-Hall, Englewood Cliffs
- Kosko B (1992) *Neural networks and fuzzy systems*. Prentice-Hall, Englewood Cliffs
- Kruse R, Meyer KD (1987) *Statistics with vague data*. Reidel, Dordrecht

- Kruse R, Gebhard J, Klawonn F (1994) Foundations of fuzzy systems. Wiley, Chichester
- Kuncheva LI (2000) Fuzzy classifier design. Springer, Heidelberg
- Li Z (2006) Fuzzy chaotic systems modeling, control, and applications. Springer, Heidelberg
- Liu B (2007) Uncertainty theory. Springer, Heidelberg
- Ma Z (2006) Fuzzy database modeling of imprecise and uncertain engineering information. Springer, Heidelberg
- Mamdani EH (1974) Application of fuzzy algorithms for the control of a simple dynamic plant. *Proc IEE* 121: 1585–1588
- Mareš M (1994) Computation over fuzzy quantities. CRC, Boca Raton
- Mendel J (2000) Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice Hall, New York
- Mendel JM, John RIB (2002) Type-2 fuzzy sets made simple. *IEEE Trans Fuzzy Syst* 10(2):117–127
- Mordeson JN, Nair PS (2001) Fuzzy mathematics. An introduction for engineers and scientists, 2nd edn. Springer, Heidelberg
- Mukaidono M (2001) Fuzzy logic for beginners. World Scientific, Singapore
- Negoita CV, Ralescu DA (1975) Application of fuzzy sets to system analysis. Birkhäuser/Halstead, Basel/New York
- Nguyen HT, Waler EA (2005) A first course in fuzzy logic, 3rd edn. CRC, Boca Raton
- Nguyen HT, Wu B (2006) Fundamentals of statistics with fuzzy data. Springer, Heidelberg
- Novák V (1989) Fuzzy sets and their applications. Hilger, Bristol/Boston
- Novák V, Perfilieva I, Močkoř J (1999) Mathematical principles of fuzzy logic. Kluwer, Boston
- Pedrycz W (1993) Fuzzy control and fuzzy systems, 2nd edn. Research Studies/Wiley, Taunton/New York
- Pedrycz W (1995) Fuzzy sets engineering. CRC, Boca Raton
- Pedrycz W (ed) (1996) Fuzzy modelling: paradigms and practice. Kluwer, Boston
- Pedrycz W, Gomide F (1998) An introduction to fuzzy sets: analysis and design. MIT Press, Cambridge
- Peeva K, Kyosev Y (2005) Fuzzy relational calculus. World Scientific, Singapore
- Petry FE (1996) Fuzzy databases. Principles and applications. Kluwer, Boston
- Piegat A (2001) Fuzzy modeling and control. Springer, Heidelberg
- Ruspini EH (1991) On the semantics of fuzzy logic. *Int J Approx Reason* 5:45–88
- Rutkowska D (2002) Neuro-fuzzy architectures and hybrid learning. Springer, Heidelberg
- Rutkowski L (2004) Flexible neuro-fuzzy systems. Structures, learning and performance evaluation. Kluwer, Dordrecht
- Seising R (2007) The fuzzification of systems. The genesis of fuzzy set theory and its initial applications – developments up to the 1970s. Springer, Heidelberg
- Smithson M (1989) Ignorance and uncertainty. Springer, Berlin
- Sousa JMC, Kaymak U (2002) Fuzzy decision making in modelling and control. World Scientific, Singapore
- Thole U, Zimmermann H-J, Zysno P (1979) On the suitability of minimum and product operator for the intersection of fuzzy sets. *Fuzzy Sets Syst* 2:167–180
- Türkşen IB (1991) Measurement of membership functions and their acquisition. *Fuzzy Sets Syst* 40:5–38
- Türkşen IB (2006) An ontological and epistemological perspective of fuzzy set theory. Elsevier, New York
- Wang Z, Klir GJ (1992) Fuzzy measure theory. Kluwer, Boston
- Wygralak M (1996) Vaguely defined objects. Representations, fuzzy sets and nonclassical cardinality theory. Kluwer, Dordrecht
- Wygralak M (2003) Cardinalities of fuzzy sets. Springer, Heidelberg
- Yager RR (1983) Quantifiers in the formulation of multiple objective decision functions. *Inf Sci* 31:107–139
- Yager RR, Filev DP (1994) Essentials of fuzzy modeling and control. Wiley, New York
- Yager RR, Kacprzyk J (eds) (1996) The ordered weighted averaging operators: theory, methodology and applications. Kluwer, Boston
- Yazici A, George R (1999) Fuzzy database modeling. Springer, Heidelberg
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1968) Probability measures of fuzzy events. *J Math Anal Appl* 23:421–427
- Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans Syst Man Cybern SMC*-2:28–44
- Zadeh LA (1975a) Fuzzy logic and approximate reasoning. *Synthese* 30:407–428
- Zadeh LA (1975b) The concept of a linguistic variable and its application to approximate reasoning. *Inf Sci* (Part I) 8:199–249, (Part II) 8:301–357, (Part III) 9:43–80
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst* 1:3–28
- Zadeh LA (1983) A computational approach to fuzzy quantifiers in natural languages. *Comput Math Appl* 9:149–184
- Zadeh LA (1985) Syllogistic reasoning in fuzzy logic and its application to usuality and reasoning with dispositions. *IEEE Trans Syst Man Cybern SMC*-15:754–763
- Zadeh LA (1986) Fuzzy probabilities. *Inf Process Manag* 20:363–372
- Zadeh LA, Kacprzyk J (eds) (1992) Fuzzy logic for the management of uncertainty. Wiley, New York
- Zadeh LA, Kacprzyk J (eds) (1999a) Computing with words in information/intelligent systems. 1 Foundations. Springer, Heidelberg

- Zadeh LA, Kacprzyk J (eds) (1999b) Computing with words in information/intelligent systems. 2 Applications. Springer, Heidelberg
- Zang H, Liu D (2006) Fuzzy modeling and fuzzy control. Birkhäuser, New York
- Zimmermann H-J (1976) Description and optimization of fuzzy systems. *Int J Gen Syst* 2:209–215
- Zimmermann H-J (1987) Fuzzy sets, decision making, and expert systems. Kluwer, Dordrecht
- Zimmermann H-J (1996) Fuzzy set theory and its applications, 3rd edn. Kluwer, Boston
- Zimmermann H-J, Zysno P (1980) Latent connectives in human decision making. *Fuzzy Sets Syst* 4:37–51

Hybrid Soft Computing Models for Systems Modeling and Control

Oscar Castillo and Patricia Melin
Division of Graduate Studies and Research,
Tijuana Institute of Technology, Tijuana, Mexico

Article Outline

Glossary
Definition of the Subject
Introduction
Genetic Algorithm for Optimization
Evolution of Fuzzy Systems
Application to Anesthesia Control
Application to the Control of the Bar and Ball System
Hierarchical Genetic Algorithms for Neural Networks
Experimental Results for Time Series Prediction
Conclusions
Future Directions
Bibliography

Glossary

Hybrid intelligent systems Intelligent Systems that are build using a combination of soft computing techniques. In particular, Soft Computing includes fuzzy logic, neural networks, genetic algorithms or hybrid approaches.

Intelligent control The application of intelligent techniques for achieving the control of non-linear plants. In particular, the use of fuzzy logic, neural networks, genetic algorithms or hybrid approaches for designing intelligent controllers.

Soft computing Soft Computing is a new area of Computer Science that deals with new intelligent methodologies that combine symbolic and numerical calculations. In particular, Soft Computing includes, at the moment,

methodologies like fuzzy logic, neural networks, genetic algorithms or hybrid approaches.

Fuzzy systems Intelligent systems that are developed based on the theory of fuzzy logic, fuzzy inference and membership functions, and fuzzy rules. Fuzzy systems are able to manage the uncertainty of the decision process of humans, and for this reason are able to mimic the expert decision process in automation applications.

Genetic algorithms Genetic algorithms are search optimization techniques that mimic natural evolution for finding solution to complex problems. In particular, genetic algorithms use operators to generate new candidate solutions based on the selection of previous good solutions.

Evolution of fuzzy systems Application of evolutionary algorithms to the optimization of number of fuzzy rules and membership functions, as well as the parameter values of the fuzzy system.

Definition of the Subject

The evolutionary design of hybrid intelligent systems using hierarchical genetic algorithms will be described in this paper. The evolutionary approach can be used for fuzzy system optimization in intelligent control. In particular, we consider the problem of optimizing the number of rules and membership functions using an evolutionary approach. The hierarchical genetic algorithm enables the optimization of the fuzzy system design for a particular application. We illustrate the approach with two cases of intelligent control. Simulation results for both applications show that we are able to find an optimal set of rules and membership functions for the fuzzy control system. We also describe the application of the evolutionary approach for the problem of designing hybrid intelligent systems in time series prediction. In this case, the goal is to design the best

predictor for complex time series. Simulation results show that the evolutionary approach optimizes the hybrid intelligent systems in time series prediction.

Introduction

The application of a Hierarchical Genetic Algorithm (HGA) for fuzzy system optimization (Man et al. 1999) will be described in this paper. In particular, we consider the problem of finding the optimal set of rules and membership functions for a specific application (Yen and Langari 1999). The HGA is used to search for this optimal set of rules and membership functions, according to the data about the problem. We consider, as an illustration, the case of a fuzzy system for intelligent control (Castillo et al. 2004).

Fuzzy systems are capable of handling complex, non-linear and sometimes mathematically intangible dynamic systems using simple solutions (Jang et al. 1997). Very often, fuzzy systems may provide a better performance than conventional non-fuzzy approaches with less development cost (Procyk and Mamdani 1979). However, to obtain an optimal set of fuzzy membership functions and rules is not an easy task (Langari 1990). It requires time, experience and skills of the designer for the tedious fuzzy tuning exercise (Valdes et al. 2005). In principle, there is no general rule or method for the fuzzy logic set-up, although a heuristic and iterative procedure for modifying the membership functions to improve performance has been proposed (Sepulveda et al. 2005). Recently, many researchers have considered a number of intelligent schemes for the task of tuning the fuzzy system (Baruch and Garrido 2005). The noticeable Neural Network (NN) approach (Jang and Sun 1995) and the Genetic Algorithm (GA) approach (Homaifar 1995) to optimize either the membership functions or rules, have become a trend for fuzzy logic system development.

The HGA approach differs from the other techniques (Davis 1991) in that it has the ability to reach an optimal set of membership functions and rules without a known fuzzy system topology (Tang et al. 1998). During the optimization

phase, the membership functions need not be fixed. Throughout the genetic operations (Holland 1975), a reduced fuzzy system including the number of membership functions and fuzzy rules will be generated (Yoshikawa et al. 1996). The HGA approach has a number of advantages:

1. An optimal and the least number of membership functions and rules are obtained,
2. no pre-fixed fuzzy structure is necessary, and,
3. simpler implementing procedures and less cost are involved.

We consider in this paper the case of automatic anesthesia control in human patients for testing the optimized fuzzy controller. We did have, as a reference, the best fuzzy controller that was developed for the automatic anesthesia control (Lozano 2004), and we consider the optimization of this controller using the HGA approach. After applying the genetic algorithm the number of fuzzy rules was reduced from 12 to 9 with a similar performance of the fuzzy controller. Of course, the parameters of the membership functions were also tuned by the genetic algorithm. We did compare the simulation results of the optimized fuzzy controllers obtained with the HGA against the best fuzzy controller that was obtained previously with expert knowledge, and control is achieved in a similar fashion. Since simulation results are similar, and the number of fuzzy rules was reduced, we can conclude that the HGA approach is a good alternative for designing fuzzy systems.

We also describe the application of the evolutionary approach for the problem of designing hybrid intelligent systems in time series prediction. In this case, the goal is to design the best predictor for complex time series. Simulation results show that the evolutionary approach optimizes the hybrid intelligent systems in time series prediction.

Genetic Algorithm for Optimization

In this paper, we used a floating-point genetic algorithm (Castillo and Melin 2003) to adjust the parameter vector θ , specifically we used the

Breeder Genetic Algorithm (BGA). The genetic algorithm is used to optimize the fuzzy system for control that will be described later. A BGA can be described by the following equation:

$$\text{BGA} = (P_g^0, N, T, \Gamma, \Delta, \text{HC}, F, \text{term}) \quad (1)$$

where: P_g^0 = initial population, N = the size of the population, T = the truncation threshold, Γ = the recombination operator, Δ = the mutation operator, HC = the hill climbing method, F = the fitness function, term = the termination criterion.

The BGA uses a selection scheme called truncation selection. The % T best individuals are selected and mated randomly until the number of offspring is equal the size of the population. The offspring generation is equal to the size of the population. The offspring generation replaces the parent population. The best individual found so far will remain in the population. Self-mating is prohibited. As a recombination operator we used "extended intermediate recombination", defined as: If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are the parents, then the successor $z = (z_1, \dots, z_n)$ is calculated by:

$$z_i = x_i + \alpha_i(y_i - x_i) \quad i = 1, \dots, n. \quad (2)$$

The mutation operator is defined as follows: A variable x_i is selected with probability p_m for mutation. The BGA normally uses $p_m = 1/n$. At least one variable will be mutated. A value out of the interval $[-\text{range}_i, \text{range}_i]$ is added to the variable. range_i defines the mutation range. It is normally set to $(0.1 \times \text{searchinterval}_i)$. searchinterval_i is the domain of definition for variable x_i . The new value z_i is computed according to

$$z_i = x_i \pm \text{range}_i \cdot \delta. \quad (3)$$

The + or – sign is chosen with probability 0.5. δ is computed from a distribution which prefers small values. This is realized as follows

$$\delta = \sum_{i=0}^{15} \alpha_i 2^i \quad \alpha_i \in [0, 1]. \quad (4)$$

Before mutation we set $\alpha_i = 0$. Then each α_i is mutated to 1 with probability $p_\delta = 1/16$. Only $\alpha_i = 1$ contributes to the sum. On the average there will be just one α_i with value 1, say α_j . Then δ is given by

$$\delta = 2^{-j}. \quad (5)$$

The standard BGA mutation operator is able to generate any point in the hypercube with center x defined by $x_i \pm \text{range}_i$. But it generates values much more often in the neighborhood of x . In the above standard setting, the mutation operator is able to locate the optimal x_i up to a precision of $\text{range}_i \cdot 2^{-150}$.

We also solved the problem with a LMS algorithm, with the purpose of have a good reference mark. To monitor the convergence rate of the LMS algorithm, we computed a short term average of the squared error $e^2(n)$ using

$$\text{ASE}(m) = \frac{1}{K} \sum_{k=n+1}^{n+K} e^2(k) \quad (6)$$

where $m = n/K = 1, 2, \dots$. The averaging interval K may be selected to be (approximately) $K = 10N$. The effect of the choice of the step size parameter Δ on the convergence rate of LMS algorithm may be observed by monitoring the $\text{ASE}(m)$.

Genetic Algorithm for Optimization

The proposed genetic algorithm is as follows:

1. We use real numbers as a genetic representation of the problem.
2. We initialize variable i with zero ($i = 0$).
3. We create an initial random population P_i in this case (P_0). Each individual of the population has n dimensions and, each coefficient of the fuzzy system corresponds to one dimension. Since we are using a fuzzy system of 25 coefficients, then our search space is of $n = 25$. The generated individuals have their coefficients in $[-3, 3]$.
4. We calculate the normalized fitness of each individual of the population using linear scaling with displacement, in the following form:

$$f'_i = f_i + \frac{1}{N} \sum |f_i| + \left| \min_i (f_i) \right| \quad \forall i.$$

5. We normalize the fitness of Each individual using:

$$F_i = \frac{f'_i}{\sum_{i=1}^N f'_i} \quad \forall i.$$

6. We sort the individuals from greater to lower fitness.
7. We use the truncated selection method, selecting the %T best individuals, for example if there are 500 individuals and, then we select $0.30 \cdot 500 = 150$ individuals.
8. We apply random crossover, to the individuals in the population (the 150 best ones) with the goal of creating a new population (of 500 individuals). Crossover with it self is not allowed, and all the individuals have to participate. To perform this operation we apply the genetic operator of extended intermediate recombination as follows:

If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are the parents, then the successors $z = (z_1, \dots, z_n)$ are calculated by, $z_i = x_i + \alpha_i(y_i - x_i)$ for $i = 1, \dots, n$ where α is a scaling factor selected randomly in the interval $[-d, 1 + d]$. In intermediate recombination $d = 0$, and for extended $d > 0$, a good choice is $d = 0.25$, which is the one that we used.

9. We apply the mutation genetic operator of BGA. In this case, we select an individual with probability $p_m = 1/n$ (where n represents the working dimension, in this case $n = 25$, which is the number of coefficients in the membership functions). The mutation operator calculates the new individuals z_i of the population in the following form: $z_i = x_i \pm \text{range}_i \delta$ we can note from this equation that we are actually adding to the original individual a value in the interval: $[-\text{range}_i, \text{range}_i]$ the range is defined as the search interval, which in this case is the domain of variable x_i , the sign

$\pm$ is selected randomly with probability of 0.5, and is calculated using the following formula,

$$\delta = \sum_{i=0}^{m-1} \alpha_i 2^{-i} \quad \alpha_i \in 0, 1.$$

Common used values in this equation are $m = 16$ and $m = 20$. Before mutation we initiate with $\alpha_i = 0$, then for each α_i we mutate to 1 with probability $p_\delta = 1/m$.

10. Let $i = i + 1$, and continue with step 4.

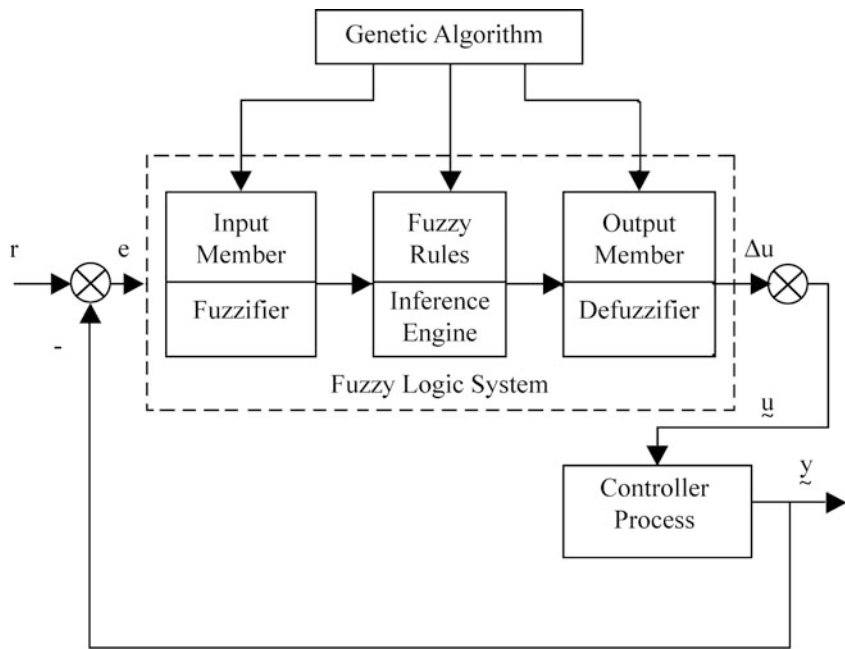
Evolution of Fuzzy Systems

Ever since the very first introduction of the fundamental concept of fuzzy logic (Zadeh 1965), its use in engineering disciplines has been widely studied. Its main attraction undoubtedly lies in the unique characteristics that fuzzy logic systems possess (Zadeh 1987). They are capable of handling complex, non-linear dynamic systems using simple solutions. Very often, fuzzy systems provide a better performance than conventional non-fuzzy approaches with less development cost (Yen and Langari 1999).

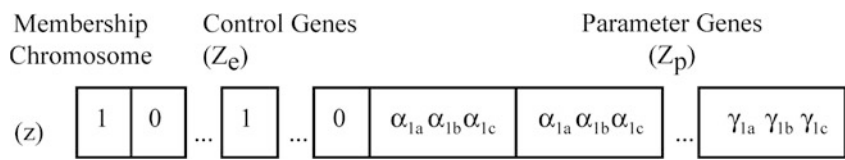
However, to obtain an optimal set of fuzzy membership functions and rules is not an easy task (Karr and Gentry 1993). It requires time, experience, and skills of the operator for the tedious fuzzy tuning exercise (Castillo and Melin 2001). In principle, there is no general rule or method for the fuzzy logic set-up (Melin and Castillo 2002). Recently, many researchers have considered a number of intelligent techniques for the task of tuning the fuzzy set.

Here, another innovative scheme is described (Valdez and Castillo 2004). This approach has the ability to reach an optimal set of membership functions and rules without a known overall fuzzy set topology. The conceptual idea of this approach is to have an automatic and intelligent scheme to tune the membership functions and rules, in which the conventional closed loop fuzzy control strategy remains unchanged, as indicated in Fig. 1.

In this case, the chromosome (Goldberg 1989) of a particular system is shown in Fig. 2. The



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 1 Genetic algorithm for a fuzzy control system



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 2 Chromosome structure for the fuzzy system

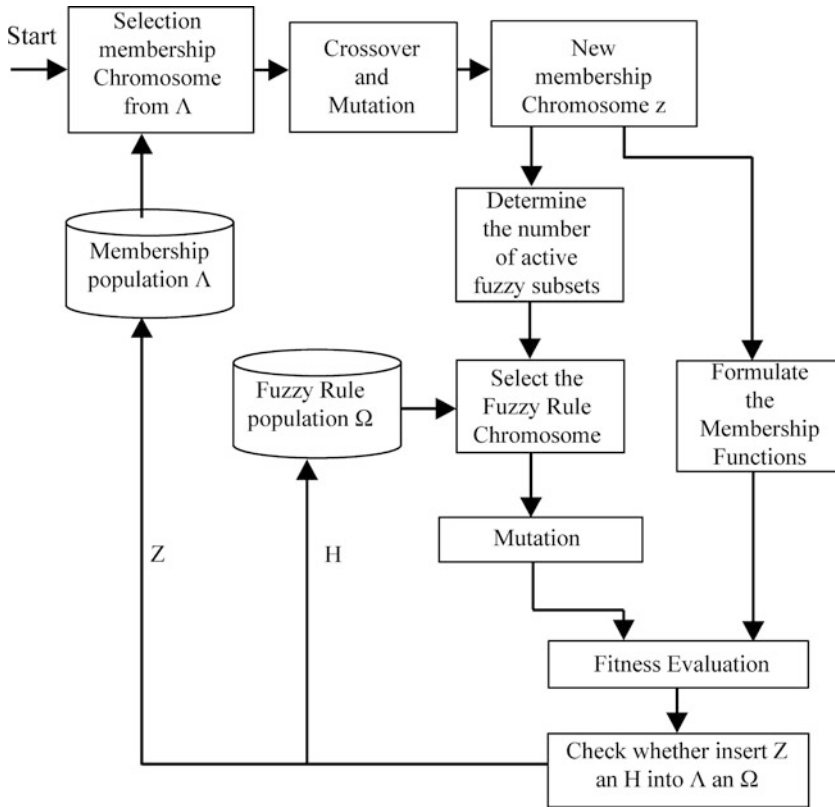
chromosome consists of two types of genes, the control genes and parameter genes. The control genes, in the form of bits, determine the membership function activation, whereas the parameter genes are in the form of real numbers to represent the membership functions.

To obtain a complete design for the fuzzy control system, an appropriate set of fuzzy rules is required to ensure system performance (Yoshikawa et al. 1996). At this point it should be stressed that the introduction of the control genes is done to govern the number of fuzzy subsystems in the system.

Once the formulation of the chromosome has been set for the fuzzy membership functions and rules, the genetic operation cycle can be

performed. This cycle of operation for the fuzzy control system optimization using a genetic algorithm is illustrated in Fig. 3.

There are two population pools, one for storing the membership chromosomes and the other for storing the fuzzy rule chromosomes. We can see this in Fig. 3 as the membership population and fuzzy rule population, respectively. Considering that there are various types of gene structure, a number of different genetic operations can be used. For the crossover operation, a one-point crossover is applied separately for both the control and parameter genes of the membership chromosomes within certain operation rates. There is no crossover operation for fuzzy rule chromosomes since only one suitable rule set can be assisted.



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 3 Genetic cycle for fuzzy system optimization

Bit mutation is applied for the control genes of the membership chromosome. Each bit of the control gene is flipped if a probability test is satisfied (a randomly generated number is smaller than a predefined rate). As for the parameter genes, which are real number represented, random mutation is applied.

The complete genetic cycle continues until some termination criteria, for example, meeting the design specification or number of generation reaching a predefined value are fulfilled.

The fitness function can be defined in this case as follows:

$$f_i = \sum |y(k) - r(k)| \quad (7)$$

where $\sum$ indicates the sum for all the data points in the training set, and $y(k)$ represents the real output of the fuzzy system and $r(k)$ is the reference

output. This fitness value measures how well the fuzzy system is approximating the real data of the problem.

Application to Anesthesia Control

We consider the case of controlling the anesthesia given to a patient as the problem for finding the optimal fuzzy system for control (Lozano 2004). The complete implementation was done in the MATLAB programming language. The fuzzy systems were build automatically by using the Fuzzy Logic Toolbox, and the genetic algorithm was coded directly in the MATLAB language. The fuzzy systems for control are the individuals used in the genetic algorithm, and these are evaluated by comparing them to the ideal control given by the experts. In other words, we compare the performance of the fuzzy systems that are

generated by the genetic algorithm, against the ideal control system given by the experts in this application. We give more details below.

Anesthesia Control Using Fuzzy Logic

The main task of the anesthetist, during and operation, is to control anesthesia concentration. In any case, anesthesia concentration can't be measured directly. For this reason, the anesthetist uses indirect information, like the heartbeat, pressure, and motor activity. The anesthesia concentration is controlled using a medicine, which can be given by a shot or by a mix of gases. We consider here the use of isoflurane, which is usually given in a concentration of 0–2% with oxygen. In Fig. 4 we show a block diagram of the controller.

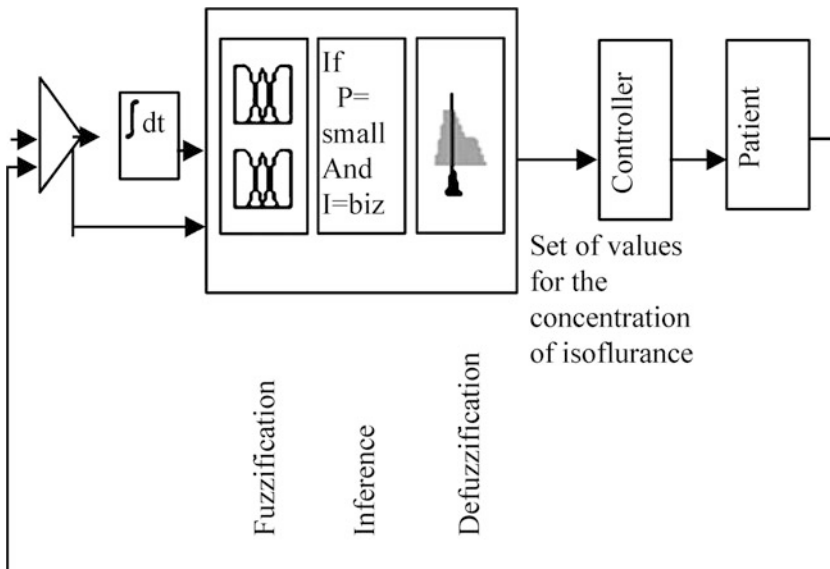
The air that is exhaled by the patient contains a specific concentration of isoflurane, and it is recirculated to the patient. As consequence, we can measure isoflurane concentration on the inhaled and exhaled air by the patient, to estimate isoflurane concentration on the patient's blood. From the control engineering point of view, the task by the anesthetist is to maintain anesthesia concentration between the high level W (threshold to wake up) and the low level E (threshold to success). These levels are difficult to be determine

in a changing environment and also are dependent on the patient's condition. For this reason, it is important to automate this anesthesia control, to perform this task more efficiently and accurately, and also to free the anesthetist from this time consuming job. The anesthetist can then concentrate in doing other task during operation of a patient.

The first automated system for anesthesia control was developed using a PID controller in the 60's. However, this system was not very successful due to the non-linear nature of the problem of anesthesia control. After this first attempt, adaptive control was proposed to automate anesthesia control, but robustness was the problem in this case. For these reasons, fuzzy logic was proposed for solving this problem. An additional advantage of fuzzy control is that we can use in the rules the same vocabulary as the medical doctors use. The fuzzy control system can also be easily interpreted by the anesthetists.

Characteristics of the Fuzzy Controller

In this section we describe the main characteristics of the fuzzy controller for anesthesia control. We will define input and output variable of the fuzzy



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 4 Architecture of the fuzzy control system

system. Also, the fuzzy rules of fuzzy controller previously designed will be described.

The fuzzy system is defined as follows:

1. Input variables: Blood pressure and Error.
2. Output variable: Isoflurance concentration.
3. Nine fuzzy if-then rules of the optimized system, which is the base for comparison.
4. 12 fuzzy if-then rules of an initial system to begin the optimization cycle of the genetic algorithm.

The linguistic values used in the fuzzy rules are the following:

- PB = Positive Big.
- PS = Positive Small.
- ZERO = zero.
- NB = Negative Big.
- NS = Negative Small.

We show below a sample set of fuzzy rules that are used in the fuzzy inference system that is represented in the genetic algorithm for optimization:

- if Blood pressure is NB and error is NB then conc_isoflurance is PS,
- if Blood pressures is PS then conc_isoflurance is NS,
- if Blood pressure is NB then conc_isoflurance is PB,
- if Blood pressure is PB then conc_isoflurance is NB,
- if Blood pressure is ZERO and error is ZERO then conc_isoflurance is ZERO,
- if Blood pressure is ZERO and error is PS then conc_isoflurance is NS,
- if Blood pressure is ZERO and error is NS then conc_isoflurance is PS,
- if error is NB then conc_isoflurance is PB,
- if error is PB then conc_isoflurance is NB,
- if error is PS then conc_isoflurance is NS,
- if Blood pressure is NS and error is ZERO then conc_isoflurance is NB,
- if Blood pressure is PS and error is ZERO then conc_isoflurance is PS.

Genetic Algorithm Specification

The general characteristics of the genetic algorithm that was used are the following:

- **NIND** = 40; % Number of individuals in each subpopulation.
- **MAXGEN** = 100; % Maximum number of generations allowed.
- **GGAP** = 0.6; % “Generational gap”, which is the percentage from the complete population of new individuals generated in each generation.
- **PRECI** = 120; % Precision of binary representations.
- **SelCh** = select('rws', Chrom, FitnV, GGAP); % Roulette wheel method for selecting the individuals participating in the genetic operations.
- **SelCh** = recomb('xovmp', SelCh, 0.7); % Multi-point crossover as recombination method for the selected individuals.
- **ObjV** = FuncionObjDifuso120_555 (Chrom, sdifuso); Objective function is given by the error between the performance of the ideal control system given by the experts and the fuzzy control system given by the genetic algorithm.

Representation of the Chromosome

In Table 1 we show the chromosome representation, which has 120 binary positions. These positions are divided in two parts, the first one indicates the number of rules of the fuzzy inference system, and the second one is divided again

Hybrid Soft Computing Models for Systems Modeling and Control, Table 1 Binary chromosome representation

Bit assigned	Representation
1–12	Which rule is active or inactive
13–21	Membership functions active or inactive of rule 1
22–30	Membership functions active or inactive of rule 2
...	Membership functions active or inactive of rule ...
112–120	Membership functions active or inactive of rule 12

into fuzzy rules to indicate which membership functions are active or inactive for the corresponding rule.

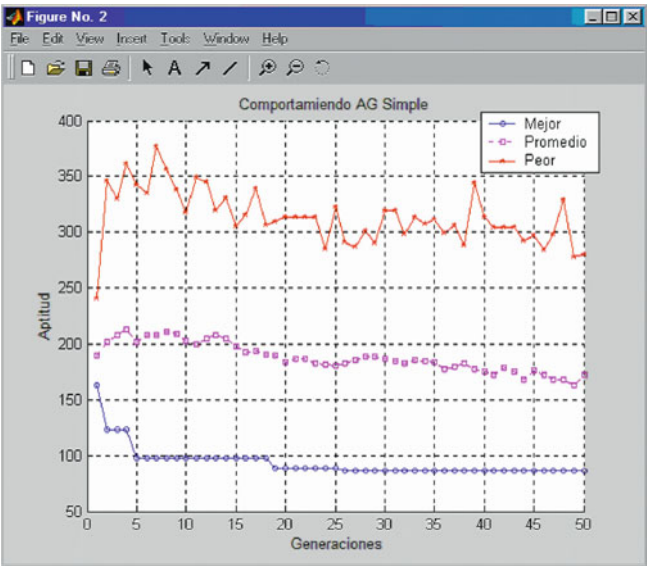
Simulation Results for the Case of Anesthesia Control

We describe in this section the simulation results that were achieved using the hierarchical genetic algorithm for the optimization of the fuzzy control system, for the case of anesthesia control. The genetic algorithm is able to evolve the topology of the fuzzy system for the particular application. We used 50 generations of 40 individuals each to

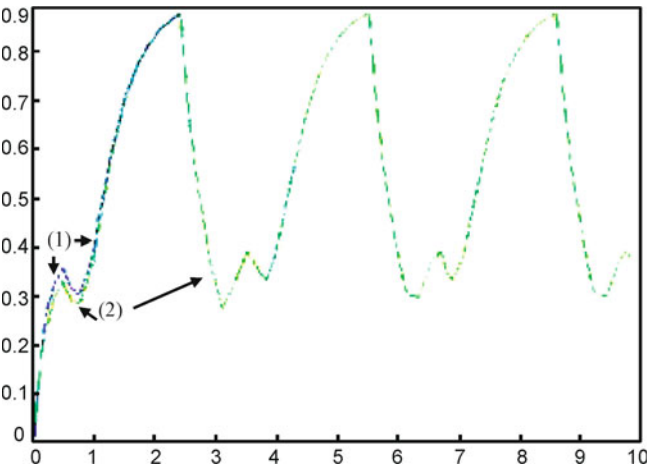
achieve the minimum error. We show in Fig. 5 the final results of the genetic algorithm, where the error has been minimized. This is the case in which only nine fuzzy rules are needed for the fuzzy controller. The value of the minimum error achieved with this particular fuzzy logic controller was of 0.0064064, which is considered a small number in this application.

In Fig. 6 we show the simulation results of the fuzzy logic controller produced by the genetic algorithm after evolution. We used a sinusoidal input signal with unit amplitude and a frequency of 2 radians/second, with a transfer function of

Hybrid Soft Computing Models for Systems Modeling and Control,
Fig. 5 Plot of the error after 50 generations of the HGA



Hybrid Soft Computing Models for Systems Modeling and Control,
Fig. 6 Comparison between outputs of the ideal controller (1) and the fuzzy controller with the HGA (2)



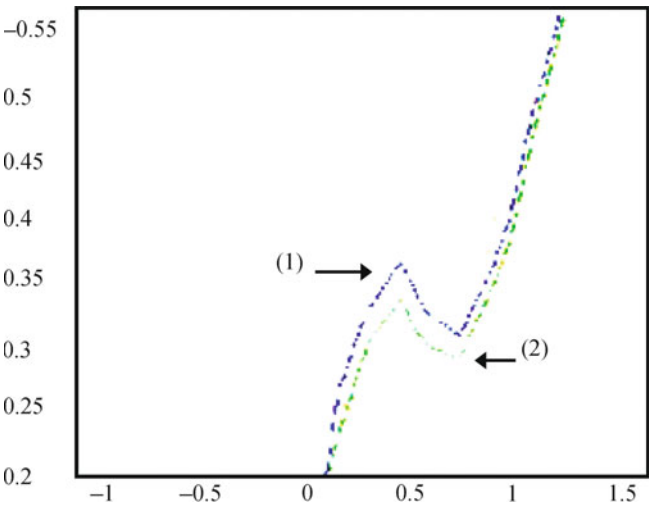
$[1/(0.5 s + 1)]$. In this figure we can appreciate the comparison of the outputs of both the ideal controller (1) and the fuzzy controller optimized by the genetic algorithm (2). From this figure it is clear that both controllers are very similar and as a consequence we can conclude that the genetic algorithm was able to optimize the performance of the fuzzy logic controller. We can also appreciate this fact more clearly in Fig. 7, where we have amplified the simulation results from Fig. 6 for a better view.

Application to the Control of the Bar and Ball System

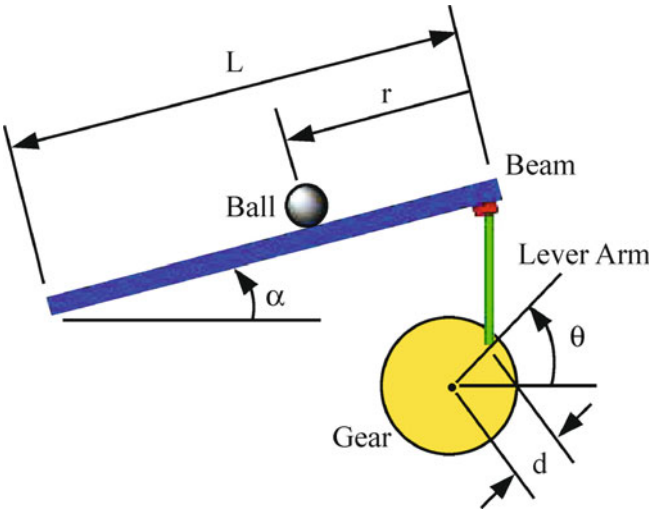
In this section, we describe the ball and beam experiment (Valdez and Castillo 2004), which was also used as a basis for testing the genetic approach of fuzzy controller optimization.

A ball is placed on a beam, see Fig. 8, where it is allowed to roll with one degree of freedom along the length of the beam. A lever arm is attached to the beam at one end and a servo gear at the other. As the servo gear turns by an angle

Hybrid Soft Computing Models for Systems Modeling and Control,
Fig. 7 Zoom in of Fig. 6 to view in more detail the difference between the controllers



Hybrid Soft Computing Models for Systems Modeling and Control,
Fig. 8 Diagram of the ball and beam system



theta, the lever changes the angle of the beam by a magnitude “alpha”. When the angle is changed from the horizontal position, gravity causes the ball to roll along the beam. A controller will be designed for this system so that the ball’s position can be manipulated.

For this problem, we will assume that the ball rolls without slipping and friction between the beam and ball is negligible. The constants for this example are defined as follows:

- M** Mass of the ball 0.11 kg
- R** Radius of the ball 0.015 m
- d** Lever arm offset 0.03 m
- g** Gravitational acceleration 9.8 m/s²
- L** Length of the beam 1.0 m
- J** Moment of inertia 9.99×10^{-6} kg m²
- r** Gravitational acceleration 9.8 m/s²
- α** Beam angle coordinate
- θ** Servo gear angle

The Lagrangian equation of motion for the ball is then given by the following equation:

$$(J/R^2 + m)r'' + mg \sin \alpha - mr(\alpha')^2 = 0. \quad (8)$$

Simulation Results with the Complete HGA

- NIND = 50; % Number of individuals in the population.
- MAXGEN = 80; % Maximum number of generations.
- GGAP = 0.8; % Generation gap.
- PRECI = 120; % Length of the Chromosome.

In this case, we use the complete HGA with the following parameters:

At the end of the genetic evolution seven fuzzy rules were obtained, which are shown in Table 2.

Simulation Results with a Method Dased Only on Mutation

In this section, we show results for the HGA method based only on mutation (no crossover was used). The parameters of the genetic algorithm are:

- NIND = 50; % Number of individuals in the population.

Hybrid Soft Computing Models for Systems Modeling and Control, Table 2 Fuzzy rules in indexed and linguistic form

Indexed rules	Linguistic rules
1 1, 1 (1): 1	If error is N and derror is N then angle is NG
1 2, 2 (1): 1	If error is N and derror is Z then angle is N
2 1, 2 (1): 1	If error is Z and derror is N then angle is N
2 2, 3 (1): 1	If error is Z and derror is Z then angle is Z
2 3, 4 (1): 1	If error is Z and derror is P then angle is P
3 2, 4 (1): 1	If error is P and derror is Z then angle is P
3 3, 5 (1): 1	If error is P and derror is P then angle is PG

Hybrid Soft Computing Models for Systems Modeling and Control, Table 3 Fuzzy rules in indexed and linguistic form

Indexed rules	Linguistic rules
1 1, 1 (1): 1	If error is N and derror is N then angle is NG
1 2, 2 (1): 1	If error is N and derror is Z then angle is N
2 1, 2 (1): 1	If error is Z and derror is N then angle is N
2 2, 3 (1): 1	If error is Z and derror is Z then angle is Z
2 3, 4 (1): 1	If error is Z and derror is P then angle is P
3 2, 4 (1): 1	If error is P and derror is Z then angle is P

- MAXGEN = 250; % Maximum number of generations.
- GGAP = 0.8; % Generation gap.
- PRECI = 120; % Length of the chromosome.

Table 3 shows the fuzzy rules obtained at the end with this type of genetic algorithm.

Simulation Results with a Method Based Only on Crossover

In this section, we show results for the HGA method based only on crossover (no mutation was used). The parameters of the genetic algorithm are the same. The fuzzy rules obtained at

the end of the genetic evolution are shown in Table 4.

Comparison of the Simulation Results

In this section we show the comparison of the fuzzy controllers that were obtained with the different types of genetic algorithms that were considered. We show in Figs. 9 and 10 the comparison of the responses of the different fuzzy controllers. From these figures we can appreciate that the different types of HGA work. However, a pseudobacterial genetic algorithm (PBGA) does not give a valid fuzzy controller. The PBGA is not described in this paper, but can be seen with detail in (Nawa and Furuhashi 1999; Salmeri et al. 1999).

Hybrid Soft Computing Models for Systems Modeling and Control, Table 4 Fuzzy rules in indexed and linguistic form

Indexed rules	Linguistic rules
1 1, 1 (1): 1	If error is N and derror is N then angle is NG
1 3, 3 (1): 1	If error is N and derror is P then angle is Z
2 1, 2 (1): 1	If error is Z and derror is N then angle is N
2 3, 4 (1): 1	If error is Z and derror is P then angle is P
3 2, 4 (1): 1	If error is P and derror is Z then angle is P
3 3, 5 (1): 1	If error is P and derror is P then angle is PG

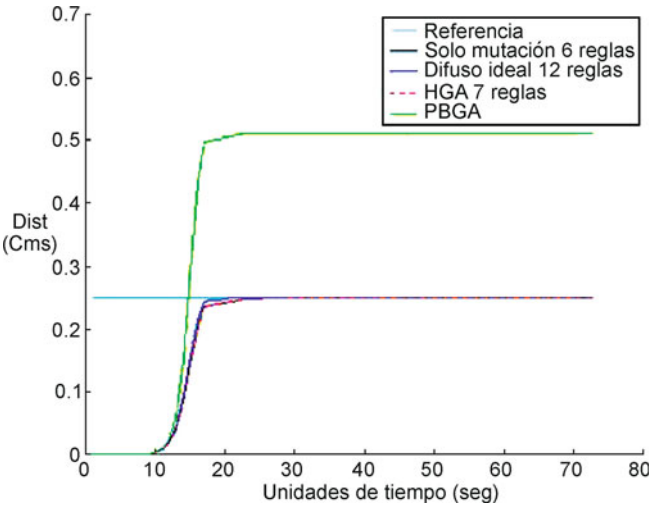
Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 9 Responses of the fuzzy controllers obtained (green = PBGA, red = HGA complete, blue = ideal fuzzy, black = HGA with only mutation)

Table 5 shows the comparison between the different methods used. This table summarizes response times of the controllers and the number of fuzzy rules of the corresponding controllers. This table also contains the information of a PID controller. We also show in Fig. 11 the behavior of the PID controller to have a basis for comparison with the other controllers.

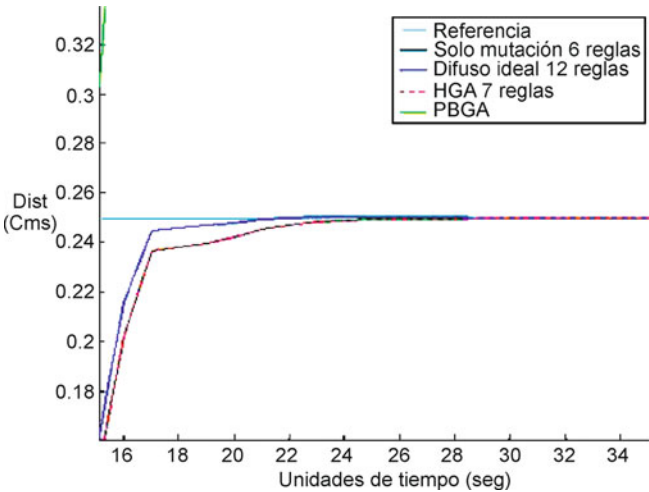
We can appreciate from Table 5 that the best result corresponds to the method using only mutation, because it achieved a lower number of fuzzy rules without losing the efficiency of the control goal. The complete HGA method also achieves control but with one more fuzzy rule. The ideal fuzzy controller (from the experts) also achieves control but with 12 fuzzy rules. The other methods do not achieve the control goal.

Hierarchical Genetic Algorithms for Neural Networks

The bottleneck problem for NN application lies within the optimization procedures that are used to obtain an optimal NN topology. Hence, the formulation of the Hierarchical Genetic Algorithm (HGA) is applied for this purpose. The HGA differs from the standard GA with a hierarchy structure in that each chromosome consists of multilevel genes. Each chromosome consists of two types of genes, i. e. control genes and



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 10 Magnification of Fig. 9



Hybrid Soft Computing Models for Systems Modeling and Control, Table 5 Comparison of the different methods

Methods	Time	Reference	Number of rules
Only mutation	12.000	0.25	6
Ideal fuzzy system (not optimized)	13.8000	0.25	12
Complete HGA	12.000	0.25	7
PBGA	Did not control	0.25	9
Only crossover	Did not control	0.25	7
Traditional (PID)	400	0.25	Transfer function

connection genes. The control genes in the form of bits, are the genes for layers and neurons for activation. The connection genes, a real value representation, are the genes for connection weightings and neuron bias.

With such a specific treatment, a structural chromosome incorporates both active and inactive genes. It should be noted that the inactive genes remain in the chromosome structure and can be carried forward for further generations. Such an inherent genetic variation in the chromosome avoids any trapping at local optima, which has the potential to cause premature convergence. Thus it maintains a balance between exploiting its accumulated knowledge and exploring the new areas of the search space. This structure also

allows larger genetic variations in chromosome while maintaining high viability by permitting multiple simultaneous genetic changes. As a result, a single change in high level genes will cause multiple changes (activation or deactivation in the whole level) in lower level genes. In the case of the traditional GA, this is only possible when a sequence of many random changes takes place. Hence the computational power is greatly improved.

The fitness function used in this work combines the information the error objective and also the information about the number of nodes as a second objective. This is shown in the following equation.

$$f(z) = \left(\frac{1}{\alpha * \text{Ranking}(\text{ObjV1}) + \beta * \text{ObjV2}} \right) * 10. \tag{9}$$

The first objective is basically the average sum of squared of errors as calculated by the predicted outputs of the MNN compared with real values of the function. This is given by the following equation.

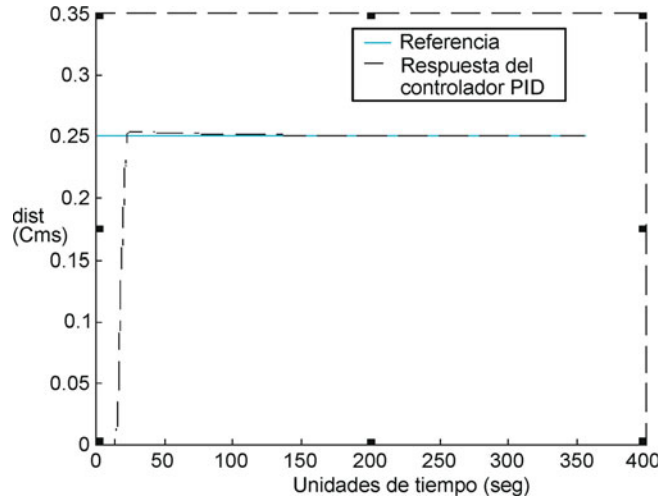
$$f_1 = \frac{1}{N} \sum_{i=1}^N (Y_i - y_i). \tag{10}$$

The parameters of the genetic algorithm for this case are as follows:

Hybrid Soft Computing Models for Systems

Modeling and Control,

Fig. 11 Response of the PID controller (*blue* = reference, *black* = response of the controller)



- Type of crossover operator: Two-point crossover.
- Crossover rate: 0.8.
- Type of mutation operator: Binary mutation.
- Mutation rate: 0.05.
- Population size per generation: 10.
- Total number of generations: 100.

The evolution of neural networks is used in the following section to optimize hybrid intelligent systems for time series prediction. In particular, in the neuro-genetic approach the evolution is used to optimize the neural network for prediction, and in the neuro-fuzzy-genetic approach the evolution is used to optimize the neuro-fuzzy system.

Experimental Results for Time Series Prediction

In this section, we illustrate the application of interval type-2 fuzzy logic to the problem of time series prediction. The Mackey-Glass time series is used to compare the results of interval type-2 fuzzy logic with the results of other intelligent methods. In particular, a comparison is made with type-1 fuzzy systems, neural networks, neuro-fuzzy systems and neuro-genetic and fuzzy-genetic approaches.

Prediction with a Neural Network

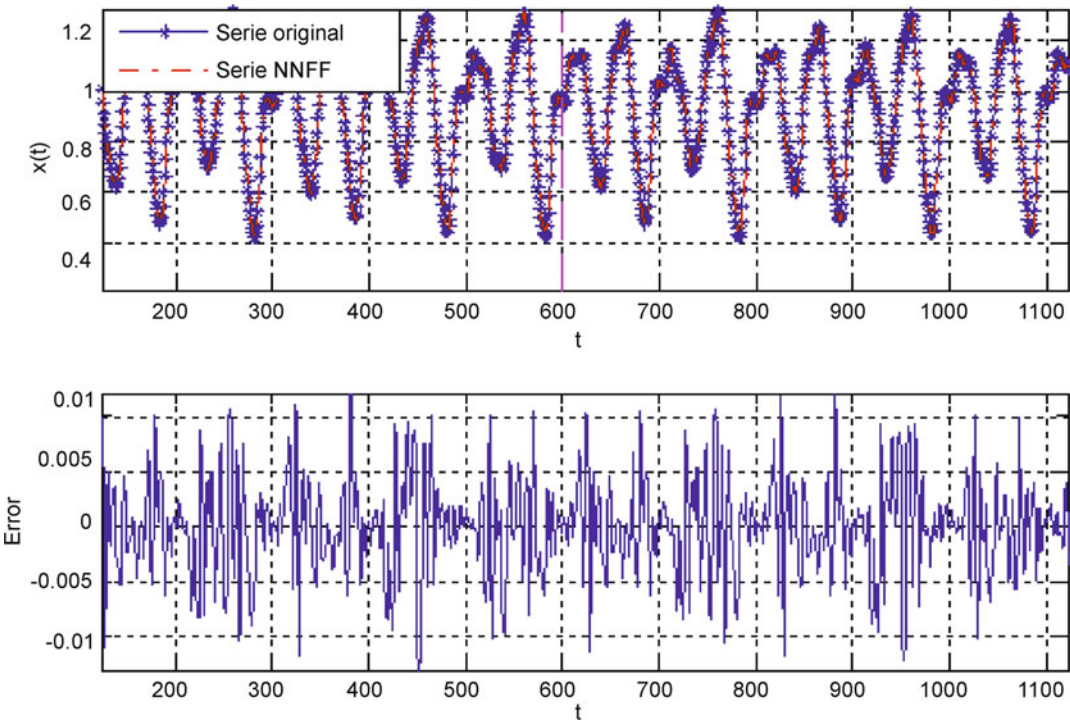
In this case, we did find a neural network model for this time series using 5 time delays and 6 periods. The architecture of the neural network used was 4–12–1 with 150 epochs, and 500 data points for training/500 data points for validation. The mean square error for prediction is 0.0043 (see Fig. 12).

Prediction with an Adaptive Neuro-Fuzzy Inference System (ANFIS)

To find the ANFIS model an analysis was made of the data, and the decision was to use a 5 time delay with 6 periods. The time series was divided into 500 data points for training and 500 data points for validation. The ANFIS model was designed with 4 inputs (2 membership functions each) and 1 output with 16 linear functions, giving a total of 16 fuzzy rules. The training was of 50 epochs and the mean squared error of prediction is of 0.0016 (see Fig. 13).

Prediction with an Interval Type-2 TSK Fuzzy Model

To find the type-2 fuzzy model an analysis was made of the data, and the decision was to use a 5 time delay with 6 periods. The time series was divided into 500 data points for training and 500 data points for validation. The model was designed with 4 inputs (with two interval type-2 (ig-bellmtype2) membership functions each) and



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 12 Forecasting the Mackey-Glass time series with a neural network. *On top* it is shown the

comparison of the prediction with the neural network (red) and the original time series (blue). *Below* it is shown the forecasting error

one output with 16 interval linear functions, giving a total of 16 fuzzy rules. The training was of 50 epochs and the mean squared error of prediction is of 0.00023 (see Fig. 14).

Prediction with a Neuro-Genetic Model

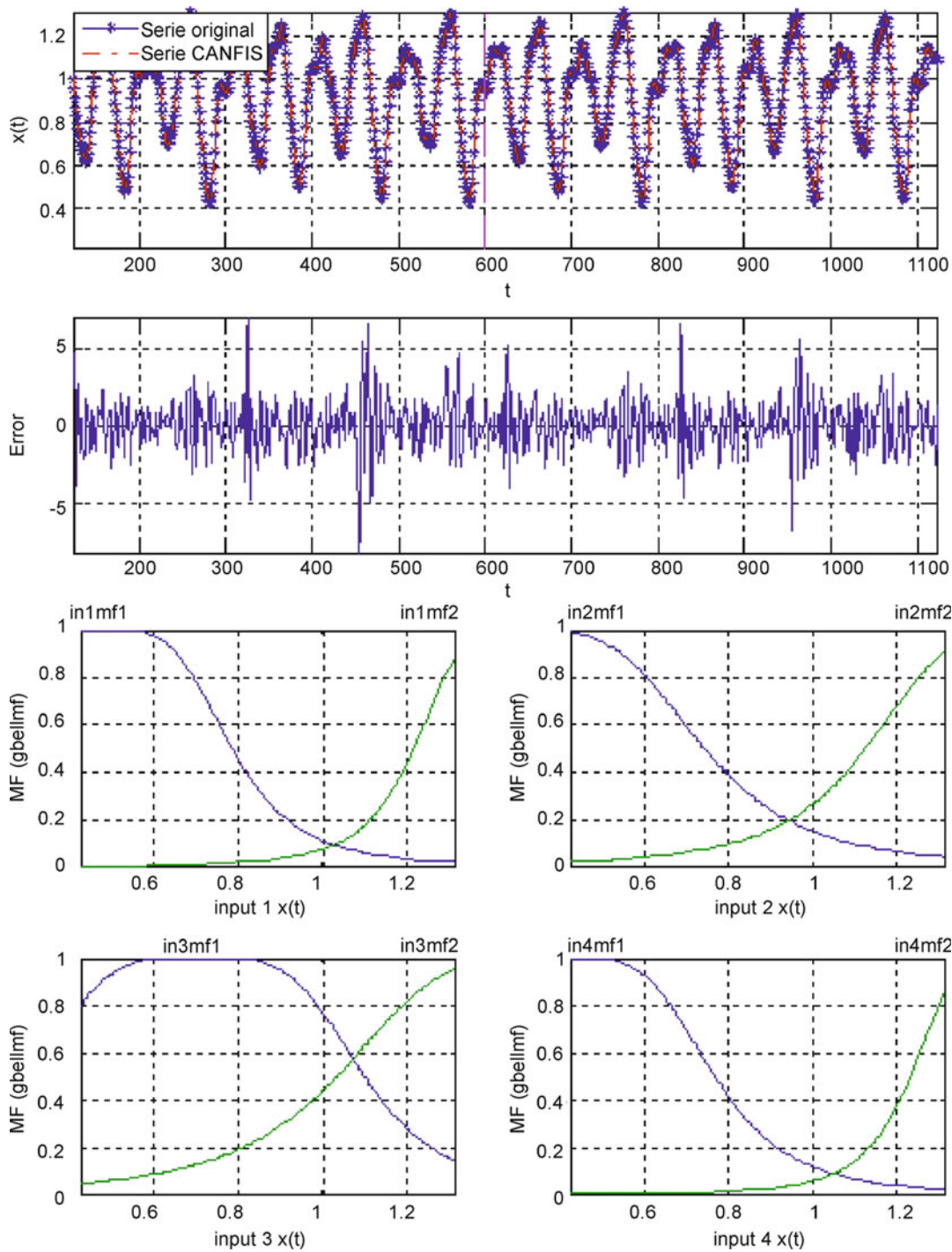
In this case, we use a genetic algorithm for training a neural network model for this time series using 5 time delays and 6 periods. The architecture of the neural network used was 4–12–1 with 200 generations, and 500 data points for training/500 data points for validation. The mean square error for prediction is 0.00064 (see Fig. 15). The genetic parameters are: population size = 30, crossover ratio = 0.75, and mutation ratio = 0.01.

Prediction with Type-1 Fuzzy Models Optimized with Genetic Algorithms

In this section, we show prediction results of type-1 fuzzy models that were optimized using

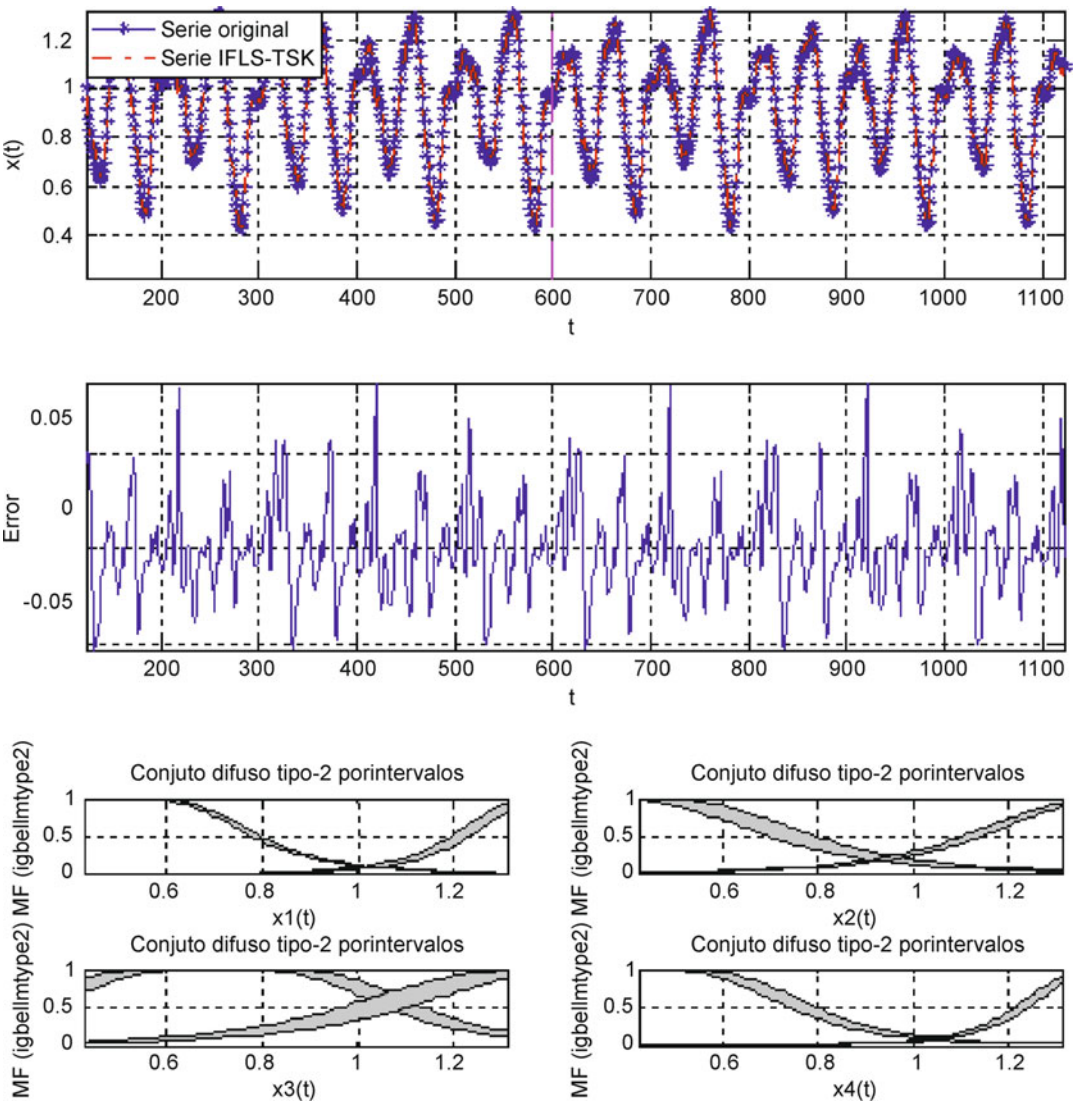
genetic algorithms. In all cases, we have 4 inputs (with 2 membership functions) and 1 output with 16 functions. The parameters of the genetic algorithms are the same as in the previous section. We show in Fig. 16 the results of a TSK model (mean squared error of 0.00647) and in Fig. 17 the results of a Mamdani model (mean squared error of 0.00692).

The Mackey-Glass time series shows chaotic behavior and for this reason has been chosen many times as a benchmark problem for prediction methods. We show in Table 6 a summary of the results using the methods mentioned previously. Based on mean squared error (RMSE) of forecasting, we can conclude that the interval type-2 fuzzy model (IT2 FLS on Table 6) is the best one to predict future values of this time series. Also, based on the training required, we can conclude that the interval type-2 fuzzy model is the best one because it only requires one epoch.



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 13 Forecasting the Mackey-Glass time series with ANFIS. The figure on *top* shows the

comparison of the predicted values and the original time series. The *following* figures show the error plot and the membership functions obtained with ANFIS



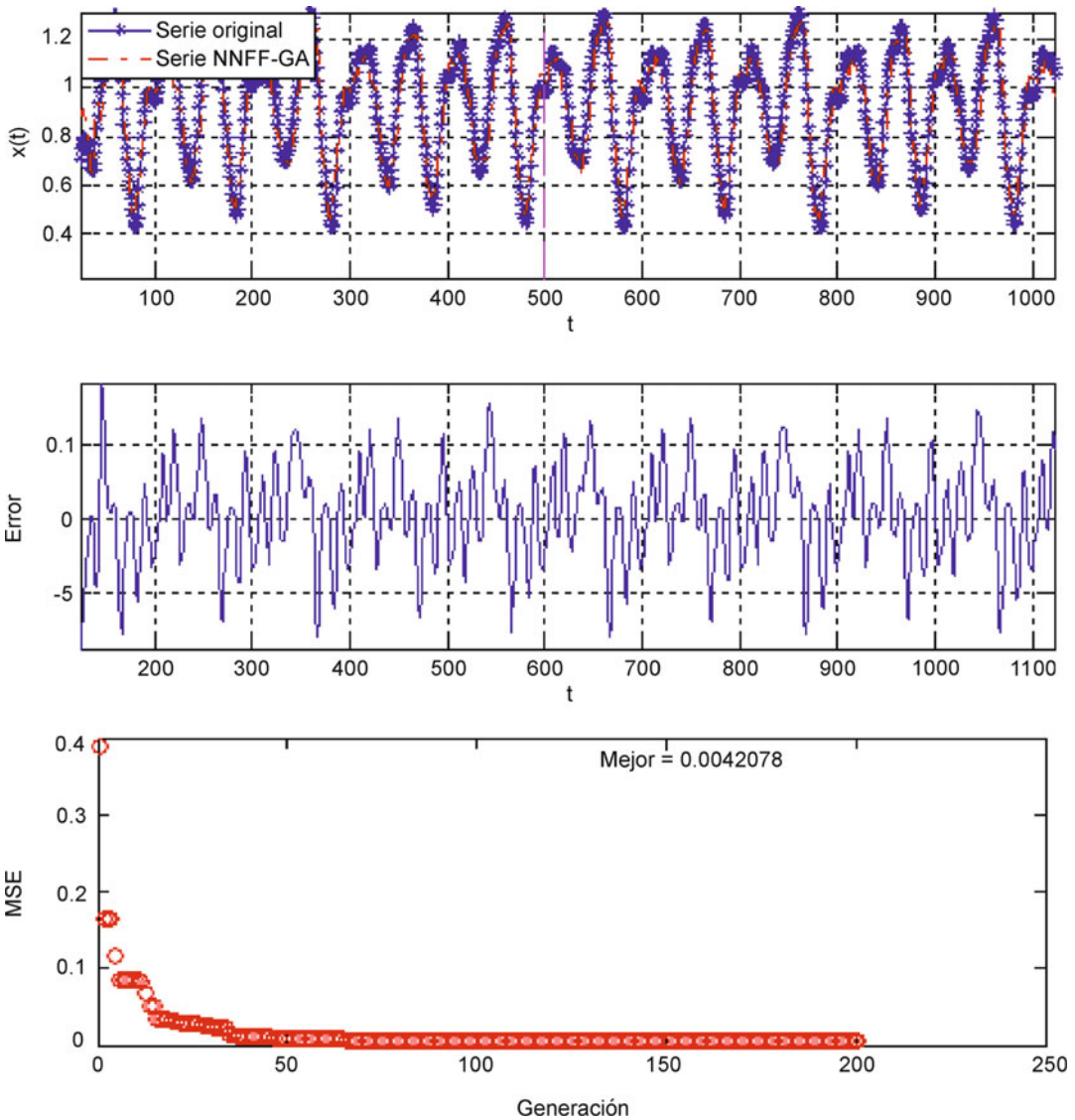
Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 14 Forecasting the Mackey-Glass time series with an interval type-2 TSK fuzzy model. The figure on *top* shows the comparison of the predicted values

and the original time series. The *following* figures show the error plot and the interval type-2 membership functions obtained

Finally, from Table 6 we can say that the use of evolution helps in the design of hybrid intelligent systems. The results of the hybrid intelligent systems shown in the last three rows (of Table 6) are very good, and this is a consequence of using genetic algorithm for optimizing the neural networks or the fuzzy systems.

Conclusions

We consider in this paper the case of automatic anesthesia control in human patients for testing the optimized fuzzy controller. We did have, as a reference, the best fuzzy controller that was developed for the automatic anesthesia control, and we consider the optimization of this controller using

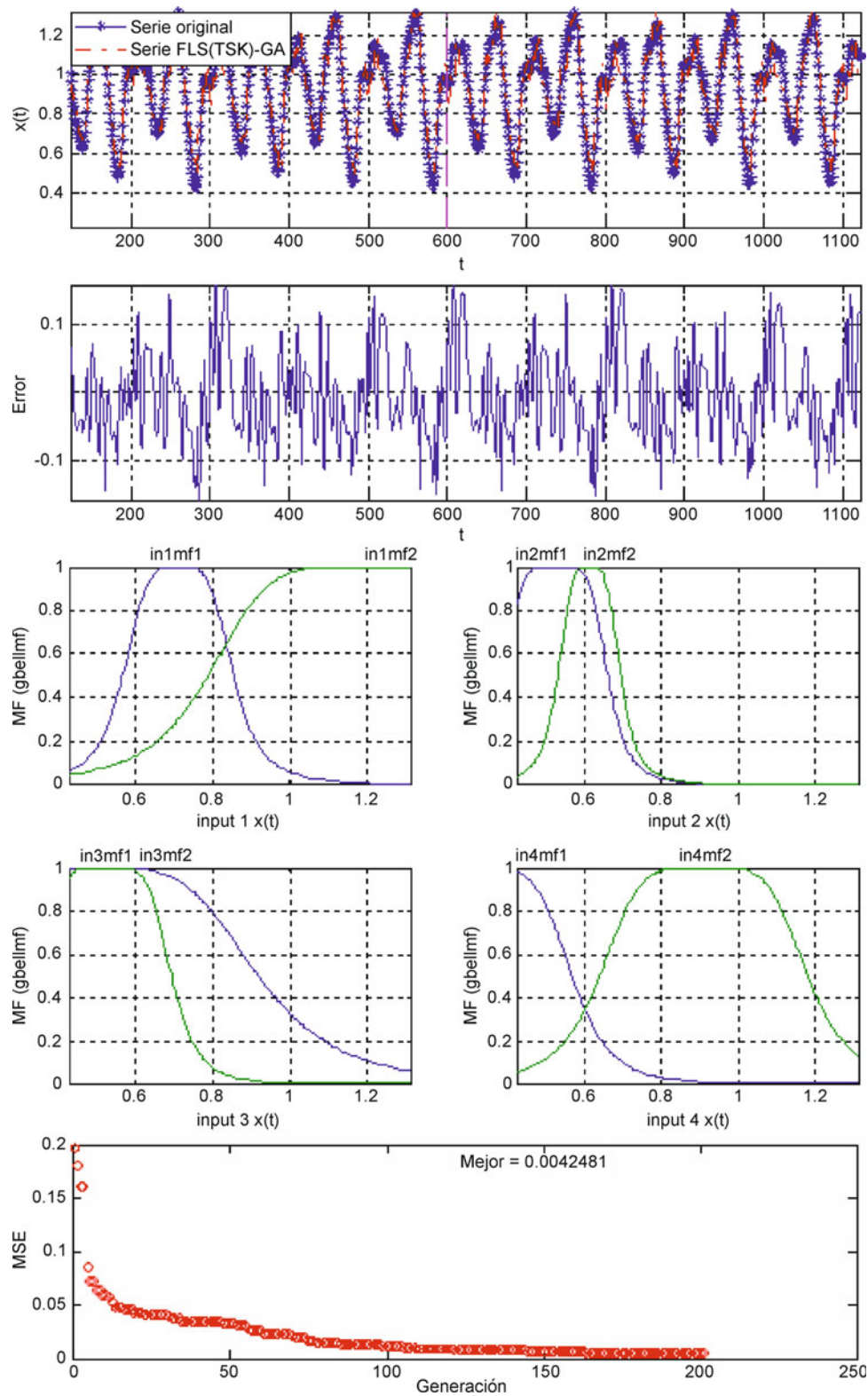


Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 15 Forecasting the Mackey-Glass time series with a neuro-genetic model. The figure on *top* shows the comparison of the predicted values and the

original time series. The figure on the *middle* shows the error plot. The figure on the *bottom* shows the evolution of the mean squared error as the generations increase up to 250

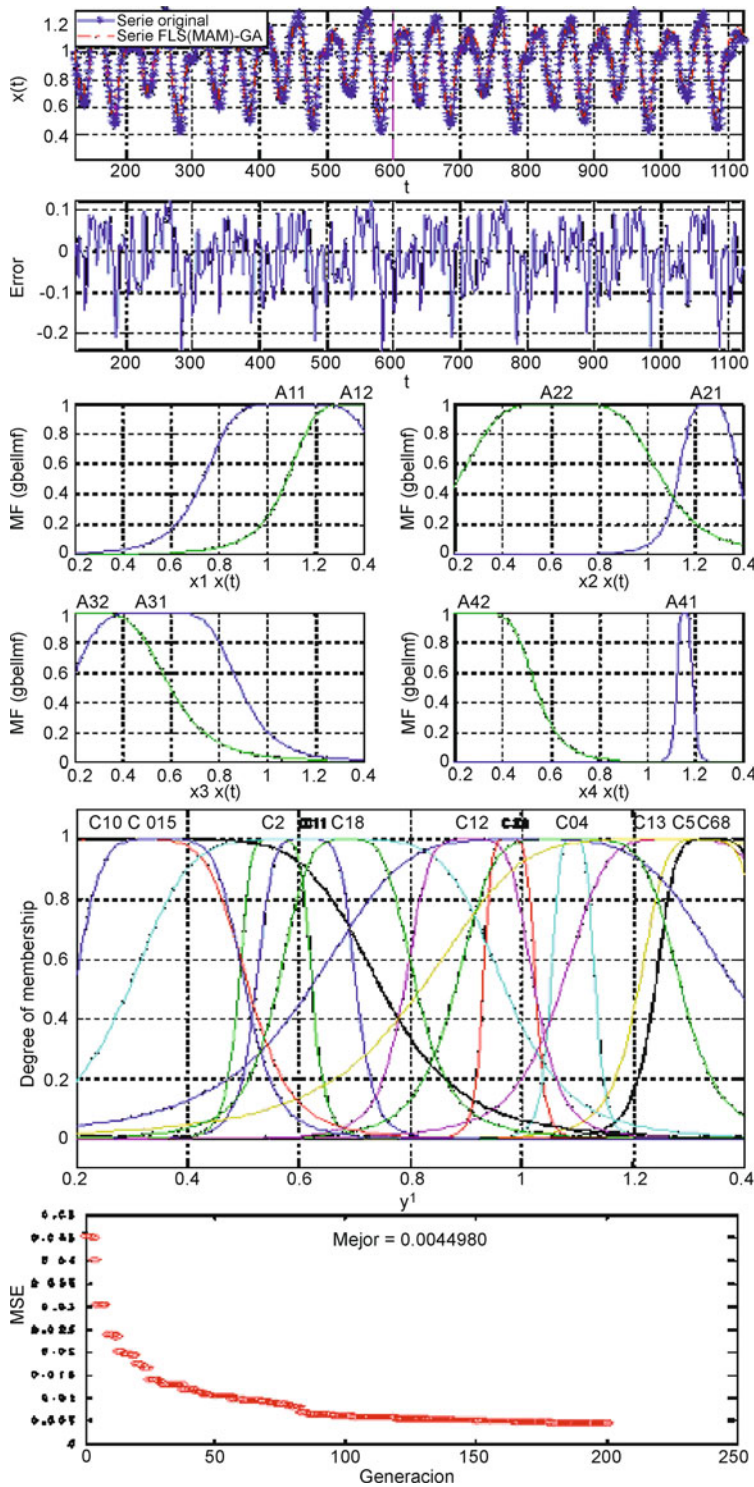
the HGA approach. After applying the genetic algorithm the number of fuzzy rules was reduced from 12 to 9 with a similar performance of the fuzzy controller. Of course, the parameters of the membership functions were also tuned by the genetic algorithm. We did compare the simulation results of the optimized fuzzy controllers obtained with the HGA against the best fuzzy controller

that was obtained previously with expert knowledge, and control is achieved in a similar fashion. Since simulation results are similar, and the number of fuzzy rules was reduced, we can conclude that the HGA approach is a good alternative for designing fuzzy systems. We also consider the case of controlling the bar and ball system, and the genetic approach was able to optimize the



Hybrid Soft Computing Models for Systems Modeling and Control, Fig. 16 Forecasting the Mackey-Glass time series with a TSK fuzzy model optimized using a genetic algorithm

Hybrid Soft Computing Models for Systems Modeling and Control,
Fig. 17 Forecasting the Mackey-Glass time series with a Mamdani fuzzy model optimized using a genetic algorithm



Hybrid Soft Computing Models for Systems Modeling and Control, Table 6 Summary of results for the mackey-glass time series

Method	RMSE	Data training/ checking	Epochs or generations
NNFF (Fig. 12)	0.00430	500/500	150
CANFIS (Fig. 13)	0.00160	500/500	50
IT2FLS (TSK) (Fig. 14)	0.00023	500/500	1
NNFF-GA (Fig. 15)	0.00064	500/500	150
FLS(TSK)- GA (Fig. 16)	0.00647	500/500	200
FLS(MAM)- GA (Fig. 17)	0.00693	500/500	200

number of rules from 12 to 6. In conclusion, the HGA approach is a good alternative in optimizing fuzzy controllers. Future work will include testing the proposed approach with the optimization of other fuzzy controllers. We also described the application of the evolutionary approach for the problem of designing hybrid intelligent systems in time series prediction. In this case, the goal is to design the best predictor for complex time series. Simulation results show that the evolutionary approach optimizes the hybrid intelligent systems in time series prediction.

Future Directions

The evolutionary approach is a good alternative in optimizing fuzzy controllers. Future work will include testing the proposed approach with the optimization of other fuzzy controllers and comparison with results of existing approaches. We also described the application of the evolutionary approach for the problem of designing hybrid intelligent systems in time series prediction. In this case, the goal is to design the best predictor for complex time series. Simulation results show that the evolutionary approach optimizes the hybrid intelligent systems in time series

prediction. In this case, future work will include applying the hybrid approach to other problems of time series prediction, and compare the results with existing approaches.

Bibliography

- Baruch S, Garrido R (2005) A direct adaptive neural control scheme with integral terms. *Int J Intell Syst* 20(2):213224
- Castillo O, Melin P (2001) Soft computing for control of non-linear dynamical systems. Springer, Heidelberg
- Castillo O, Melin P (2003) Soft computing and fractal theory for intelligent manufacturing. Springer, Heidelberg
- Castillo O, Huesca G, Valdez F (2004) Evolutionary computing for fuzzy system optimization in intelligent control. *Proceedings of IC-AI'04, Las Vegas*, vol 1. CSREA Press, Las Vegas, pp 98–104
- Davis L (1991) Handbook of genetic algorithms. Van Nostrand Reinhold, New York
- Goldberg D (ed) (1989) Genetic algorithms in search, optimization and machine learning. Addison Wesley, Reading
- Holland J (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
- Homaifar MCE (1995) Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Trans Fuzzy Syst* 3:129–139
- Jang JSR, Sun CT (1995) Neurofuzzy fuzzy modeling and control. *Proc IEEE* 83:378–406
- Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing, a computational approach to learning and machine intelligence. Prentice Hall, Upper Saddle River
- Karr CL, Gentry EJ (1993) Fuzzy control of pH using genetic algorithms. *IEEE Trans Fuzzy Systems* 1: 46–53
- Langari R (1990) A framework for analysis and synthesis of fuzzy linguistic control systems. Ph D thesis, University of California, Berkeley
- Lozano (2004) Optimización de un sistema de control difuso por medio de algoritmos genéticos jerárquicos. Thesis, Department of Computer Science, Tijuana Institute of Technology
- Man KF, Tang KS, Kwong S (1999) Genetic algorithms: concepts and designs. Springer, London
- Melin P, Castillo O (2002) Modelling, simulation and control of non-linear dynamical systems. Taylor and Francis, London
- Nawa NE, Furuhashi T (1999) Fuzzy system parameters discovery by bacterial evolutionary algorithm. *IEEE Trans Fuzzy Syst* 7:608–616
- Procyk TJ, Mamdani EM (1979) A linguistic self-organizing process controller. *Automatica* 15(1):15–30
- Salmeri M, Re M, Petrongari E, Cardarilli GC (1999) A novel bacterial algorithm to extract the rule

- base from a training set. Technical Report, Department of Electronic Engineering, University of Rome
- Sepulveda R, Castillo O, Melin P, Montiel O, Rodriguez-Diaz A (2005) Handling uncertainty in controllers using type-2 fuzzy logic. *J Intell Syst* 14:237–262
- Tang KS, Man KF, Liu ZF, Kwong S (1998) Minimal fuzzy memberships and rules using hierarchical genetic algorithms. *IEEE Trans Ind Electron* 45(1):142–150
- Vachtsevanos G, Farinwata S (1996) Fuzzy logic control. In: Patyra MJ, Mlynek DM (eds) *A systematic design and performance assessment methodology*. Wiley, New York
- Valdes M, Gomez-Skarmeta AF, Botia JA (2005) Toward a framework for the specification of hybrid fuzzy modeling. *Int J Intell Syst* 20(2):225–252
- Valdez F, Castillo O (2004) Comparative study of evolutionary computing methods for fuzzy system optimization in intelligent control. *Proc IS-IC'04*. Tijuana, México, pp 1–5
- Yen J, Langari R (1999) *Fuzzy logic: intelligence, control and information*. Prentice Hall, Upper Saddle River
- Yoshikawa T, Furuhashi T, Uchikawa Y (1996) Emergence of effective fuzzy rules for controlling mobile robots using DNA coding method. *Proc ICEC'96*. Nagoya, Japan, pp 581–586
- Zadeh L (1965) Fuzzy sets. *J Inf. Control* 8:338–353
- Zadeh L (1987) Fuzzy sets and applications. In: Yager RR, Ovchinnikov S, Tong RM, Nguyen HT (eds) *Selected papers*. Wiley, New York

Neuro-fuzzy Systems

Leszek Rutkowski¹, Krzysztof Cpałka^{1,2},
Robert Nowicki¹, Agata Pokropinska³ and
Rafal Scherer¹

¹Departament of Computer Engineering,
Czestochowa University of Technology,
Czestochowa, Poland

²Department of Artificial Intelligence,
Academy of Humanities and Economics,
Lodz, Poland

³Institute of Mathematics and Computer
Science, Jan Dlugosz University, Czestochowa,
Poland

Article Outline

Glossary
Definition of the Subject
Introduction
Fuzzy Reasoning
Description of Fuzzy Inference Systems
Logical-Type Neuro-fuzzy Systems
Mamdani-Type Neuro-fuzzy Systems
Simplified Neuro-fuzzy Systems
Takagi–Sugeno Neuro-fuzzy Systems
Neuro-fuzzy Systems with Weights
Neuro-fuzzy Systems for Pattern Classification
Learning
Criteria Isolines Method
Future Directions
Bibliography

Glossary

Neuro-fuzzy systems Fusion of fuzzy logic and neural networks with the ability to automated adaptation to training data and knowledge interpretability.

Fuzzy reasoning Reasoning on the basis of fuzzy premises and fuzzy rules inferring fuzzy conclusions.

Definition of the Subject

Neuro-fuzzy systems (NFS) are part of soft computing concept. They are a synergistic fusion of fuzzy logic and neural networks with the ability to automate adaptation to training data and knowledge interpretability. Thus neuro-fuzzy systems incorporate two goals of machine learning which are difficult to meet at the same time: transparency of the rule base and high accuracy obtained thanks to learning from data. There are various types of NFS which differ in the form of fuzzy rules and inference methods. They have a very broad range of applications in life, science, economics, manufacturing, medicine etc. NFS can replace neural networks in all applications as most of them are universal approximators. Therefore they can perform well in tasks of classification, prediction, approximation and control, similarly to neural networks. Neural networks work as black boxes (usually it is not possible to tell where the knowledge is stored) and can not use prior knowledge. NFS can use almost the same learning methods and achieve the same accuracy as neural networks yet the knowledge in the form of fuzzy rules is easily interpretable for humans. On the other hand, fuzzy systems (not neuro-fuzzy) require manual tuning of their parameters and fuzzy reasoning in such systems is more trouble some and computationally demanding than simplified one, used in NFS.

Introduction

Over the last decade fuzzy sets and fuzzy logic introduced in 1965 by Lotfi Zadeh (1965) have been used in a wide range of problem domains, including process control, image processing, pattern recognition and classification, management, economics and decision making. Specific applications include washing-machine automation, cam-corder focusing, TV color tuning, automobile

transmissions and subway operations (Hirota 1993). We have also been witnessing a rapid development in the area of neural networks (see e. g. Tadeusiewicz 1993; Tadeusiewicz 1998; Żurada 1992). Both fuzzy systems and neural networks, along with probabilistic methods (Aliev and Aliev 2001; Eubank 1999; Pagan and Ullah 1999), evolutionary algorithms (Fogel 1995; Michalewicz 1992), rough sets (Pawlak 1982, 1991) and uncertain variables (Bubnicki 2001, 2002a, b), constitute a consortium of soft computing techniques (Aliev and Aliev 2001; Kasabov 1996; Kecman 2001). These techniques are often used in combination. For example, fuzzy inference systems are frequently converted into connectionist structures called neuro-fuzzy systems which exhibit advantages of neural networks and fuzzy systems. In literature various neuro-fuzzy systems have been developed (see e. g. Chen and Linkens 2001; Chuang et al. 2001; Corcoran and Sen 1994; Czogała 2000; Duan and Chung 2001; Fuller 2000; Gaweda and Żurada 2000; Gaweda and Żurada 2001; Jang and Sun 1995; Jang et al. 1997; Juang and Lin 1998; Kasabov 1996; Lee et al. 1994, 1996; Lin 1994; Lin and Lee 1991, 1997; Lin and Lu 1995; Lin and Cunningham III 1995; Mouzouris and Mendel 1997; Nauck and Kruse 1996; Nie and Linkens 1995; Pedrycz 1992; Roubos and Setnes 2001; Rutkowska 2002; Rutkowski and Cpałka 2000, 2003; Wang 1994). They combine the natural language description of fuzzy systems and the learning properties of neural networks. Some of them are known in literature under short names such as ANFIS (Jang 1993), ANNBFIS (Czogała 2000), DENFIS (Kasabov 2002), FALCON (Lin and Lee 1991), GARIC (Berenji and Khedkar 1992), NEFCLASS (Nauck et al. 1997), NEFPROX (Nauck and Kruse 1999; Nauck et al. 1997), SANFIS (Wang and Lee 2002) and others. The most popular designs of neuro-fuzzy structures fall into one of the following categories, depending on the connective between the antecedent and the consequent in fuzzy rules:

- (i) Takagi–Sugeno method – consequents are functions of inputs,

- (ii) Mamdani-type reasoning method – consequents and antecedents are related by the min operator or generally by a t-norm,
- (iii) Logical-type reasoning method – consequents and antecedents are related by fuzzy implications, e. g. binary, Lukasiewicz, Zadeh and others.

It should be noted that most applications are dominated by the Mamdani-type fuzzy reasoning. However, it was emphasized by Yager (1990, 1992) that “no formal reason exists for the preponderant use of the Mamdani method in fuzzy logic control as opposed to the logical method other than inertia”. In this work we will outline all three types of systems. Moreover we will compare several variants of these systems in terms of efficiency for an exemplary problem.

Fuzzy Reasoning

In this section we present the idea of fuzzy reasoning with various fuzzy implications which will be useful for construction of neuro-fuzzy systems developed in the next sections. The basic rule of inference in classical logic is modus ponens. The compositional rule of inference describes a composition of a fuzzy set and a fuzzy relation. Fuzzy rule

$$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B \quad (1)$$

is represented by a fuzzy relation R . Having given input linguistic value A' , we can infer an output fuzzy set B' by the composition of the fuzzy set A' and the relation R . The generalized modus ponens is the extension of the conventional modus ponens tautology, to allow partial fulfillment of the premises:

Premise	$x \text{ is } A'$
Implication	$\text{IF } x \text{ is } A \text{ THEN } y \text{ is } B$
Conclusion	$y \text{ is } B'$

where A, B, A', B' are fuzzy sets, x and y are linguistic variables. Applying the compositional rule of inference (Kacprzyk 1997), we obtain

$$B' = A' \circ R = A' \circ (A \rightarrow B) \quad (2)$$

and

$$\begin{aligned} \mu_{B'}(y) &= \mu_{A' \circ R}(y) \\ &= \sup_{x \in X} \left\{ \mu_{A'}(x)^T * \mu_R(x, y) \right\}. \end{aligned} \quad (3)$$

The problem is to determine the membership function of the fuzzy relation described by

$$\mu_R(x, y) = \mu_{A \rightarrow B}(x, y) \quad (4)$$

based on the knowledge of $\mu_A(x)$ and $\mu_B(y)$. We denote

$$\mu_{A \rightarrow B}(x, y) = I(\mu_A(x), \mu_B(y)), \quad (5)$$

where $I(\cdot) = I_{\text{fuzzy}}(\cdot)$ is a fuzzy implication (logical approach to fuzzy reasoning, see subsection “[Logical Approach to Fuzzy Inference Reasoning](#)”) given in Definition 1 or a t-norm (Mamdani approach to fuzzy reasoning, see subsection “[Mamdani Approach to Fuzzy Inference Reasoning](#)”).

Logical Approach to Fuzzy Inference Reasoning

Definition 1 (see Fodor 1991) A fuzzy implication is a function $I = I_{\text{fuzzy}}: [0, 1]^2 \rightarrow [0, 1]$ satisfying the following conditions:

- (I1) if $a_1 \leq a_3$, then $I(a_1, a_2) \geq I(a_3, a_2)$,
for all $a_1, a_2, a_3 \in [0, 1]$,
- (I2) If $a_2 \leq a_3$, then $I(a_1, a_2) \leq I(a_1, a_3)$,
For all $a_1, a_2, a_3 \in [0, 1]$,
- (I3) $I(0, a_2) = 1$, for all $a_2 \in [0, 1]$
(falsity implies anything),
- (I4) $I(a_1, 1) = 1$, for all $a_1 \in [0, 1]$
(anything implies tautology),
- (I5) $I(1, 0) = 0$ (Booleanity).

Selected fuzzy implications satisfying all or some of the above conditions are listed in Table 1.

In this table, Implications 1–4 are examples of an S-implication associated with a t-conorm

Neuro-fuzzy Systems, Table 1 Fuzzy implications

No	Name	Implication $I_{\text{fuzzy}}(a, b)$
1	Kleene–Dienes (binary)	$\max\{1 - a, b\}$
2	Lukasiewicz	$\min\{1, 1 - a + b\}$
3	Reichenbach	$1 - a + a \cdot b$
4	Fodor	$\begin{cases} 1 & \text{if } a \leq b \\ \max\{1 - a, b\} & \text{if } a > b \end{cases}$
5	Rescher	$\begin{cases} 1 & \text{if } a \leq b \\ 0 & \text{if } a > b \end{cases}$
6	Goguen	$\begin{cases} 1 & \text{if } a = 0 \\ \min\left\{1, \frac{b}{a}\right\} & \text{if } a > 0 \end{cases}$
7	Gödel	$\begin{cases} 1 & \text{if } a \leq b \\ b & \text{if } a > b \end{cases}$
8	Yager	$\begin{cases} 1 & \text{if } a = 0 \\ b^a & \text{if } a > 0 \end{cases}$
9	Zadeh	$\max\{\min\{a, b\}, 1 - a\}$
10	Willmott	$\min\left\{\max\{1 - a, b\}, \max\{a, 1 - b, \min\{1 - a, b\}\}\right\}$
11	Dubois–Prade	$\begin{cases} 1 - a & \text{if } b = 0 \\ b & \text{if } a = 1 \\ 1 & \text{if otherwise} \end{cases}$

$$I(a, b) = S\{1 - a, b\}. \quad (6)$$

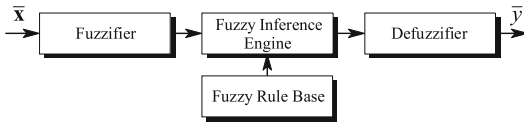
Neuro-fuzzy systems based on fuzzy implications given in Table 1 are called logical systems. Implications (6) and (7) belong to a group of R-implications associated with the t-norm T and given by

$$I(a, b) = \sup_z \{z | T\{a, z\} \leq b\}, \quad a, b \in [0, 1]. \quad (7)$$

The Zadeh implication belongs to a group of Q-implications given by

$$I(a, b) = S\{N(a), T\{a, b\}\}, \quad a, b \in [0, 1]. \quad (8)$$

It is easy to verify that S-implications and R-implications satisfy all the conditions of Definition 1. However, the Zadeh implication violates conditions I1 and I4, whereas the Willmott implication violates conditions I1, I3 and I4.



Neuro-fuzzy Systems, Fig. 1 Fuzzy inference system

Mamdani Approach to Fuzzy Inference Reasoning

In practice we frequently use Mamdani-type operators given by

$$I(a, b) = \min\{a, b\}, \quad a, b \in [0, 1], \quad (9)$$

$$I(a, b) = a \cdot b, \quad a, b \in [0, 1] \quad (10)$$

or generally

$$I(a, b) = T\{a, b\}, \quad a, b \in [0, 1]. \quad (11)$$

It should be noted that operators (9)–(11) do not satisfy conditions of Definition 1. Operators (9)–(11) are called “correlation functions” describing strength of connections between antecedents and consequences (see Rutkowski 2004).

Description of Fuzzy Inference Systems

In this description, we consider multi-input-single-output fuzzy system mapping $\mathbf{X} \rightarrow \mathbf{Y}$, where $\mathbf{X} \subset \mathbf{R}^n$ and $\mathbf{Y} \subset \mathbf{R}$ (see Fig. 1). The system is composed of a fuzzifier, a fuzzy rule base, a fuzzy inference engine and a defuzzifier. The fuzzifier performs a mapping from the observed crisp input space $\mathbf{X} \subset \mathbf{R}^n$ to a fuzzy set defined in $\mathbf{X}$. The most commonly used fuzzifier is the singleton fuzzifier, which maps $\bar{\mathbf{x}} = [\bar{x}_1, \dots, \bar{x}_n] \in \mathbf{X}$ into a fuzzy set $A' \subseteq \mathbf{X}$ characterized by the membership function

$$\mu_{A'}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \bar{\mathbf{x}} \\ 0 & \text{if } \mathbf{x} \neq \bar{\mathbf{x}} \end{cases} \quad (12)$$

The fuzzy rule base consists of a collection of N fuzzy IF-THEN rules, aggregated by the disjunction or the conjunction, in the form

$$R^{(k)} : \begin{cases} \text{IF} & x_1 \text{ is } A_1^k \quad \text{AND} \\ & x_2 \text{ is } A_2^k \quad \text{AND} \dots \\ & x_n \text{ is } A_n^k \\ \text{THEN} & y \text{ is } B^k \end{cases} \quad (13)$$

or

$$R^{(k)} : \text{IF } \mathbf{x} \text{ is } A^k \quad \text{THEN } y \text{ is } B^k, \quad (14)$$

where $\mathbf{x} = [x_1, \dots, x_n] \in \mathbf{X}$, $y \in \mathbf{Y}$, $A^k = A_1^k \times A_2^k \times \dots \times A_n^k$, $A_1^k, A_2^k, \dots, A_n^k$ are fuzzy sets characterized by membership functions $\mu_{A_i^k}(x_i)$, $i = 1, \dots, n$, $k = 1, \dots, N$, whereas B^k are fuzzy sets characterized by membership functions $\mu_{B^k}(y)$, $k = 1, \dots, N$. The firing strength of the k th rule, $k = 1, \dots, N$, is defined by

$$\tau_k(\bar{\mathbf{x}}) = \prod_{i=1}^n \left\{ \mu_{A_i^k}(\bar{x}_i) \right\} = \mu_{A^k}(\bar{\mathbf{x}}). \quad (15)$$

The fuzzy inference engine determines the mapping from the fuzzy sets in the input space $\mathbf{X}$ to the fuzzy sets in the output space $\mathbf{Y}$. Each of N rules (14) determines a fuzzy set $\bar{B}^k \subseteq \mathbf{Y}$ given by the compositional rule of inference

$$\bar{B}^k = A'^k \circ (A^k \rightarrow B^k), \quad (16)$$

where $A^k = A_1^k \times A_2^k \times \dots \times A_n^k$. Fuzzy sets $\bar{B}^k$ are characterized by membership functions expressed by the sup-star composition

$$\mu_{\bar{B}^k}(y) = \sup_{\mathbf{x} \in \mathbf{X}} \left\{ \mu_{A'}(\mathbf{x}) * \mu_{A_1^k \times \dots \times A_n^k \rightarrow B^k}(\mathbf{x}, y) \right\}, \quad (17)$$

where $*$ can be any operator in the class of t-norms. It is easily seen that for a crisp input $\bar{\mathbf{x}} = [\bar{x}_1, \dots, \bar{x}_n] \in \mathbf{X}$, i. e., the singleton fuzzifier (12), formula (17) becomes

$$\begin{aligned} \mu_{\bar{B}^k}(y) &= \mu_{A_1^k \times \dots \times A_n^k \rightarrow B^k}(\bar{\mathbf{x}}, y) \\ &= \mu_{A^k \rightarrow B^k}(\bar{\mathbf{x}}, y) \\ &= I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(y)), \end{aligned} \quad (18)$$

where $I(\cdot)$ is correlation between antecedents and consequents given by (11), in the case of Mamdani approach, or fuzzy implication $I_{\text{fuzzy}}(\cdot)$,

in the case of logical approach (examples of fuzzy implications are given in Table 1). The aggregation operator, applied in order to obtain the fuzzy set B' based on fuzzy sets $\bar{B}^k$, is the t-norm or t-conorm operator, depending on the type of the fuzzy inference. As a result of the fuzzy reasoning we obtain the fuzzy set B' .

The defuzzifier performs a mapping from the fuzzy set B' to a crisp point $\bar{y}$ in $\mathbf{Y} \subset \mathbf{R}$. The COA (center of area) method is defined by the following formula:

$$\bar{y} = \frac{\int_{\mathbf{Y}} y \cdot \mu_{B'}(y) dy}{\int_{\mathbf{Y}} \mu_{B'}(y) dy} \quad (19)$$

or by

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \mu_{B'}(\bar{y}^r)}{\sum_{r=1}^N \mu_{B'}(\bar{y}^r)}, \quad (20)$$

in the discrete form, where $\bar{y}^r$ are centers of the membership functions $\mu_{B'}(y)$, i. e., for $r = 1, \dots, N$

$$\mu_{B'}(\bar{y}^r) = \max_{y \in \mathbf{Y}} \{\mu_{B^r}(y)\}. \quad (21)$$

Logical-Type Neuro-fuzzy Systems

In this approach, function $I(\cdot)$ given by 18 is a fuzzy implication (see Table 1), i. e.

$$I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)) = I_{\text{fuzzy}}(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r)). \quad (22)$$

When we use the logical model, the aggregation is carried out by

$$B' = \bigcap_{k=1}^N \bar{B}^k. \quad (23)$$

The aggregated output fuzzy set $B' \subseteq \mathbf{Y}$ is computed by the use of a t-norm and is given by

$$\begin{aligned} \mu_{B'}(\bar{y}^r) &= \bigwedge_{k=1}^N \{\mu_{\bar{B}^k}(\bar{y}^r)\} \\ &= \bigwedge_{k=1}^N \{I_{\text{fuzzy}}(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{y}^r))\} \end{aligned} \quad (24)$$

and formula (20) becomes

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \prod_{k=1}^N \left\{ I_{\text{fuzzy}} \left(\prod_{i=1}^n \{\mu_{A_i^k}(\bar{x}_i)\}, \mu_{B^k}(\bar{y}^r) \right) \right\}}{\sum_{r=1}^N \prod_{k=1}^N \left\{ I_{\text{fuzzy}} \left(\prod_{i=1}^n \{\mu_{A_i^k}(\bar{x}_i)\}, \mu_{B^k}(\bar{y}^r) \right) \right\}} \quad (25)$$

Example 1 Using Gaussian membership functions

$$\mu_{A_i^r}(x_i) = \exp \left[- \left(\frac{x_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right], \quad (26)$$

$$\mu_{B^r}(y) = \exp \left[- \left(\frac{y - \bar{y}^r}{\sigma^r} \right)^2 \right], \quad (27)$$

dependency (25), Łukasiewicz-type S-implication (see Table 1), and the min-type aggregation of conclusions, we obtain the following description of the neuro-fuzzy system

$$\begin{aligned} \bar{y} &= \frac{\sum_{r=1}^N \bar{y}^r \cdot \min_{k=1, \dots, N} \left\{ \min \left[1, 1 - \frac{n}{i=1} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right) \right. \right. \\ &\quad \left. \left. + \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right] \right\}}{\sum_{r=1}^N \min_{k=1, \dots, N} \left\{ \min \left[1, 1 - \frac{n}{i=1} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right) \right. \right. \\ &\quad \left. \left. + \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right] \right\}} \end{aligned} \quad (28)$$

Example 2 Using Gaussian membership functions (26) and (27), dependency (25), Kleene–Dienes-type S-implications (see Table 1), and the min-type aggregation of conclusions, we obtain the following description of the neuro-fuzzy system:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \min_{k=1, \dots, N} \left\{ \max \left[1 - \frac{n}{T} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right), \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right] \right\}}{\sum_{r=1}^N \cdot \min_{k=1, \dots, N} \left\{ \max \left[1 - \frac{n}{T} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right), \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right] \right\}} \quad (29)$$

Example 3 Using Gaussian membership functions (26) and (27), dependency (25), Reichenbach-type S-implications (see Table 1), and the min-type aggregation of conclusions, we obtain the following description of the neuro-fuzzy system:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \min_{k=1, \dots, N} \left\{ 1 - \frac{n}{T} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right) \cdot \left(1 - \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right) \right\}}{\sum_{r=1}^N \cdot \min_{k=1, \dots, N} \left\{ 1 - \frac{n}{T} \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right) \cdot \left(1 - \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right) \right\}} \quad (30)$$

Mamdani-Type Neuro-fuzzy Systems

In this approach, function $I(\cdot)$ given by (18) is a t-norm (e. g. minimum or algebraic), i. e.

$$I(\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{\mathbf{y}}^r)) = T\{\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{\mathbf{y}}^r)\}. \quad (31)$$

In case of the Mamdani approach, the aggregation is carried out by

$$B' = \bigcup_{k=1}^N \bar{B}^k. \quad (32)$$

The aggregated output fuzzy set $B' \subseteq \mathbf{Y}$ is computed by the use of a t-conorm and is given by

$$\begin{aligned} \mu_{B'}(\bar{\mathbf{y}}^r) &= \bigvee_{k=1}^N \left\{ \mu_{\bar{B}^k}(\bar{\mathbf{y}}^r) \right\} \\ &= \bigvee_{k=1}^N \left\{ T\{\mu_{A^k}(\bar{\mathbf{x}}), \mu_{B^k}(\bar{\mathbf{y}}^r)\} \right\}. \end{aligned} \quad (33)$$

Consequently, formula (20) takes the form

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \bigvee_{k=1}^N \left\{ T\left\{ \frac{n}{T} \left\{ \mu_{A_i^k}(\bar{x}_i) \right\}, \mu_{B^k}(\bar{\mathbf{y}}^r) \right\} \right\}}{\sum_{r=1}^N \bigvee_{k=1}^N \left\{ T\left\{ \frac{n}{T} \left\{ \mu_{A_i^k}(\bar{x}_i) \right\}, \mu_{B^k}(\bar{\mathbf{y}}^r) \right\} \right\}}. \quad (34)$$

Obviously, the t-norms used to connect the antecedents in the rules and to connect the antecedents and consequents (correlation function) do not have to be the same. Besides, they can be chosen as differentiable functions as e. g. Yager families (Klement et al. 2000).

Remark 1 Another Mamdani-type neuro-fuzzy system can be derived using so-called “height defuzzifier”

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \mu_{\bar{B}^r}(\bar{\mathbf{y}}^r)}{\sum_{r=1}^N \mu_{\bar{B}^r}(\bar{\mathbf{y}}^r)}. \quad (35)$$

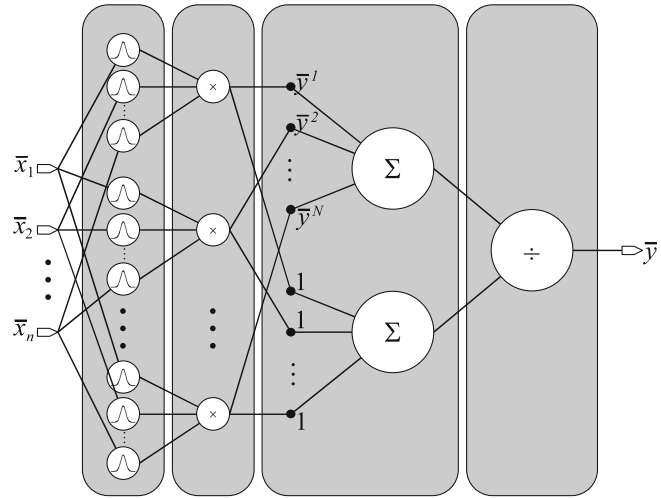
For normal fuzzy sets B^r , combining (35), (18) and (31), we obtain

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \frac{n}{T} \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}{\sum_{r=1}^N \frac{n}{T} \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}, \quad (36)$$

and for Gaussian functions (26) and (27) and the product t-norm we get

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \left[\prod_{i=1}^n \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right) \right]}{\sum_{r=1}^N \left[\prod_{i=1}^n \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right) \right]}. \quad (37)$$

It should be noted that (37) is a well-known formula derived by Wang (1994) and used for

Neuro-fuzzy Systems,**Fig. 2** Architecture of the neuro-fuzzy system with Mamdani approach described by (37)

tasks of modeling and control. It is interesting that formula (37) can be also derived from defuzzifier (20) if assumption (40) holds (see section “Fuzzy Reasoning”). The architecture of the system (37) is depicted in Fig. 2.

Example 4 Using Gaussian membership functions (26) and (27), dependency (34), min-type Mamdani rule, and max-type aggregation of conclusions, we obtain the following description of the neuro-fuzzy system

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \max_{k=1, \dots, N} \left\{ \min \left(\prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \cdot \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right\}}{\sum_{r=1}^N \max_{k=1, \dots, N} \left\{ \min \left(\prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \cdot \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right\}}. \quad (38)$$

Example 5 Using Gaussian membership functions (26) and (27), dependency (34), product-type Mamdani rule (known as the Larsen rule), and the max-type aggregation of conclusions, we obtain the following description of the neuro-fuzzy system:

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \cdot \max_{k=1, \dots, N} \left\{ \prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \cdot \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right\}}{\sum_{r=1}^N \max_{k=1, \dots, N} \left\{ \prod_{i=1}^n \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \cdot \exp \left[- \left(\frac{\bar{y}^r - \bar{y}^k}{\sigma^k} \right)^2 \right] \right\}}. \quad (39)$$

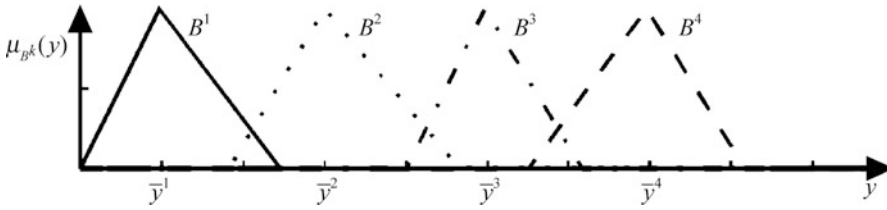
Simplified Neuro-fuzzy Systems

The structures presented so far are quite complex and complicated. One possibility of their simplification is to assume, that membership functions of consequent fuzzy sets are sufficiently distant from each other, so that the following assumption holds

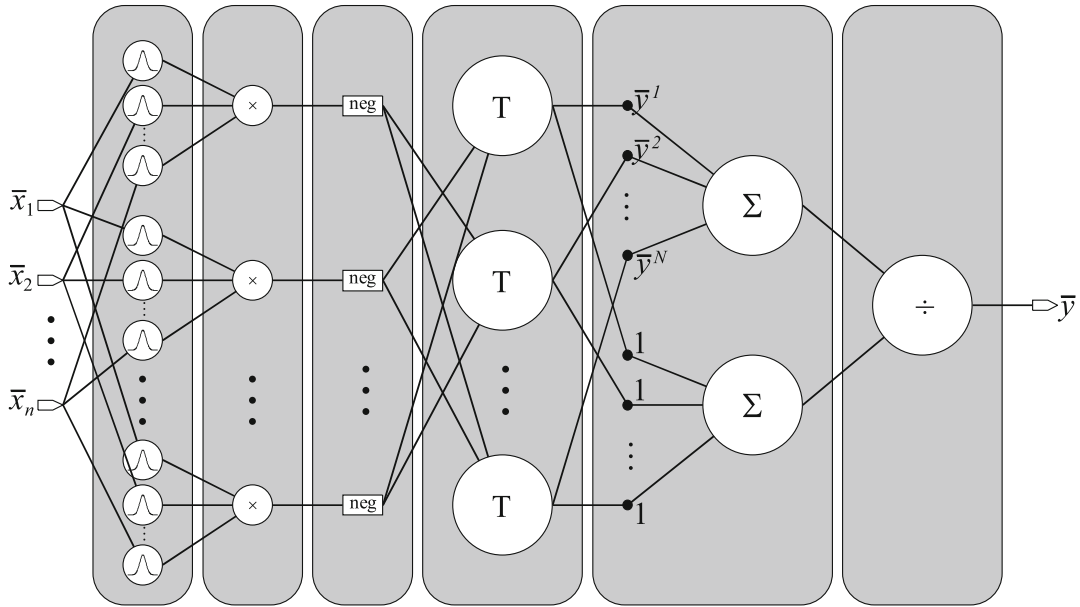
$$\mu_{B^k}(\bar{y}^r) \approx 0, \quad (40)$$

for $k, r = 1, \dots, N$ and $k \neq r$. Figure 3 shows an example of such fuzzy sets.

It is easy to verify that under assumption (40) the Mamdani-type neuro-fuzzy structure given by (34), derived from defuzzifier (20), takes the simplified form which is the same as the structure (36) derived from defuzzifier (35). The



Neuro-fuzzy Systems, Fig. 3 Exemplary fuzzy sets satisfying assumption (40)



Neuro-fuzzy Systems, Fig. 4 Architecture of the logical neuro-fuzzy system with Lukasiewicz implication described by (41)

logical neuro-fuzzy system (25) can also be simplified if condition (40) holds, e. g. the Lukasiewicz neuro-fuzzy system (28) takes the form

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \prod_{k=1}^n \left\{ 1 - \frac{n}{T} \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right\} \right\}}{\sum_{r=1}^N \prod_{k=1}^n \left\{ 1 - \frac{n}{T} \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right\} \right\}} \quad (41)$$

Let us observe that simplified neuro-fuzzy systems are characterized by three parameters to be found in the process of learning: $\bar{x}_i^k$, σ_i^k and $\bar{y}^k$ for $i = 1, \dots, n$ and $k = 1, \dots, N$. In the case of not simplified systems, see (28)–(30), we have to additionally determine parameters σ^k , $k = 1, \dots, N$, of the output membership functions. The architecture of the system (41) is depicted in Fig. 4.

Takagi–SugenoNeuro-fuzzy Systems

In the fuzzy Takagi–Sugeno-type model (Takagi and Sugeno 1985), the rules have fuzzy character

only in the IF part, whereas in the THEN part, there are functional dependencies

$$\begin{aligned} R^{(r)} : & \text{ IF } \mathbf{x} \text{ is } A^r \\ & \text{ THEN } y_r = f^{(r)}(x_1, x_2, \dots, x_n). \end{aligned} \quad (42)$$

If we assume that the input of the fuzzy system is signal $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$, then in order to obtain the output signal $\bar{y}$ of the system, first we will determine

$$T\left(\mu_{A_1^r}(\bar{x}_1), \mu_{A_2^r}(\bar{x}_2), \dots, \mu_{A_n^r}(\bar{x}_n)\right), \quad r = 1, \dots, N. \quad (43)$$

The next step is to compute

$$\bar{y}_r = f^{(r)}(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n), \quad r = 1, \dots, N. \quad (44)$$

The output signal of the fuzzy Takagi–Sugeno system is a normalized weighted sum of particular inputs $\bar{y}_1, \dots, \bar{y}_N$, i. e.

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}_r \prod_{i=1}^n \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}{\sum_{r=1}^N \prod_{i=1}^n \left\{ \mu_{A_i^r}(\bar{x}_i) \right\}}. \quad (45)$$

In the following part of this subchapter, we will consider the Takagi–Sugeno systems with linear dependencies in consequents of the base of rules, i. e.

$$\begin{aligned} R^{(r)} : & \text{ IF } \mathbf{x} \text{ is } A^r \\ & \text{ THEN } y_r = c_0^{(r)} + c_1^{(r)} x_1 + \dots + c_n^{(r)} x_n \end{aligned} \quad (46)$$

for $r = 1, \dots, N$. It should be noted that if $c_i^{(r)} = 0$, $i = 1, \dots, n$, then system (45) is reduced to a simplified Mamdani system given by formula (36), and then $c_0^{(r)} = \bar{y}^r$, $r = 1, \dots, N$.

Example 6 The Takagi–Sugeno system with Gaussian membership functions, linear model (46) and product t-norm, takes the following form

$$\bar{y} = \frac{\sum_{r=1}^N \left[\prod_{i=1}^n \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right) \cdot \left(c_0^{(r)} + c_1^{(r)} x_1 + \dots + c_n^{(r)} x_n \right) \right]}{\sum_{r=1}^N \left[\prod_{i=1}^n \left(\exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^r}{\sigma_i^r} \right)^2 \right] \right) \right]}. \quad (47)$$

Neuro-fuzzy Systems with Weights

We can distinguish fuzzy systems based on the following linguistic models:

- (i) Systems based on linguistic model given by (14).
- (ii) Systems based on linguistic model given by (14), and additionally with weights $w_k^{\text{agr}} \in [0, 1]$, $k = 1, \dots, N$, describing the importance of each rule:

$$R^{(k)} : \left[\begin{array}{l} \text{IF } x_1 \text{ is } A_1^k \text{ AND } \dots \\ \text{AND } x_n \text{ is } A_n^k \\ \text{THEN } y \text{ is } B^k \end{array} \right] (w_k^{\text{agr}}). \quad (48)$$

- (iii) Systems based on linguistic model given by (14), with weights $w_k^{\text{agr}} \in [0, 1]$, $k = 1, \dots, N$, describing the importance of each rule and weights $w_{i,k}^{\tau} \in [0, 1]$, $k = 1, \dots, N$, $i = 1, \dots, n$, describing the importance of antecedents:

$$R^{(k)} : \left[\begin{array}{l} \text{IF } x_1 \text{ is } A_1^k(w_{1,k}^{\tau}) \text{ AND } \dots \\ \text{AND } x_n \text{ is } A_n^k(w_{n,k}^{\tau}) \\ \text{THEN } y \text{ is } B^k \end{array} \right] (w_k^{\text{agr}}). \quad (49)$$

In order to incorporate the weights in models (48) and (49) into description of neuro-fuzzy systems, we can use the following formula (see Rutkowski 2004; Rutkowski and Cpałka 2003):

$$T^*\{a_1, \dots, a_n; w_1, \dots, w_n\} = \bigwedge_{i=1}^n \{1 - w_i(1 - a_i)\}, \quad (50)$$

where $T\{\cdot\}$ is any t-norm, and the weights meet the condition $w_i \in [0, 1]$, $i = 1, \dots, n$. Observe that if we assume that $w_i = 1$, $i = 1, \dots, n$, then function $T^*\{\cdot\}$ is reduced to t-norm $T\{\cdot\}$. Analogously, we define

$$S^*\{a_1, \dots, a_n; w_1, \dots, w_n\} = \bigvee_{i=1}^n \{w_i a_i\}. \quad (51)$$

where S is any t co-norm.

Example 7 Simplified neuro-fuzzy system (41) with Lukasiewicz fuzzy implication and weights describing the importance of each rule takes the form

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \bigwedge_{k=1}^n \left\{ 1 - \bigwedge_{i=1}^n \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right\}, w_k \right\}}{\sum_{r=1}^N \bigwedge_{k=1}^n \left\{ 1 - \bigwedge_{i=1}^n \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] \right\}, w_k \right\}}. \quad (52)$$

Example 8 Simplified neuro-fuzzy system (41) with Lukasiewicz fuzzy implication, weights describing the importance of each rule and weights describing the importance of antecedents takes the form

$$\bar{y} = \frac{\sum_{r=1}^N \bar{y}^r \bigwedge_{k=1}^n \left\{ 1 - \bigwedge_{i=1}^n \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] w_{i,k} \right\} w_k \right\}}{\sum_{r=1}^N \bigwedge_{k=1}^n \left\{ 1 - \bigwedge_{i=1}^n \left\{ \exp \left[- \left(\frac{\bar{x}_i - \bar{x}_i^k}{\sigma_i^k} \right)^2 \right] w_{i,k} \right\} w_k \right\}}. \quad (53)$$

Neuro-fuzzy Systems for Pattern Classification

We will explain how to modify linguistic model (13) to solve multi-classification problems. Let $[x_1, \dots, x_n]$ be the vector of features of an object v . Let $\Omega = \{\omega_1, \dots, \omega_M\}$ be a set of classes. The knowledge is represented by a set of N rules in the form

$$R^{(k)} : \begin{cases} \text{IF} & x_1 \text{ is } A_1^k \text{ AND} \\ & x_2 \text{ is } A_2^k \text{ AND} \dots \\ & x_n \text{ is } A_n^k \\ \text{THEN} & v \in \omega_1(z_1^k), \\ & v \in \omega_2(z_2^k), \dots \\ & v \in \omega_M(z_M^k), \end{cases}, \quad (54)$$

where z_j^k , $j = 1, \dots, M$, $k = 1, \dots, N$, are interpreted as “support” for class ω_j given by rule $R^{(k)}$. We will now redefine description (54). Let us introduce vector $\mathbf{z} = [z_1, \dots, z_M]$, where z_j , $j = 1, \dots, M$, is the “support” for class ω_j given by all M rules. We can scale the support values to the interval $[0, 1]$, so that z_j is the membership degree of an object v to class ω_j , according to all M rules. The rules are represented by

$$R^{(k)} : \begin{cases} \text{IF} & x_1 \text{ is } A_1^k \text{ AND} \\ & x_2 \text{ is } A_2^k \text{ AND} \dots \\ & x_n \text{ is } A_n^k \\ \text{THEN} & z_1 \text{ is } B_1^k \text{ AND} \\ & z_2 \text{ is } B_2^k \text{ AND} \dots \\ & z_M \text{ is } B_M^k \end{cases}, \quad (55)$$

and formula (20) adopted for classification takes the form

$$\bar{z}_j = \frac{\sum_{k=1}^N \bar{z}_j^k \mu_{B_j^k}(\bar{z}_j^k)}{\sum_{k=1}^N \mu_{B_j^k}(\bar{z}_j^k)}, \quad (56)$$

where $\bar{z}_j^k$ are centers of fuzzy sets B_j^k , $j = 1, \dots, M$, $k = 1, \dots, N$.

Learning

Fuzzy and neuro-fuzzy system knowledge is expressed in the form of fuzzy rules (14). The learning process tries to pick the right number of rules and location of linguistic values (fuzzy sets defined in linguistic variables) to act like an expert or device performing given task. Traditional fuzzy systems are devised to work on the basis of knowledge formulated by an expert in the form of fuzzy rules in nearly natural language. The expert is not always able to formulate his or her knowledge or he or she is not available at all. In such cases we can rely on knowledge acquisition. The data acquired are used to supervised learning and then the learning process becomes similar to that known from neural networks, especially gradient methods. To determine all parameters of neuro-fuzzy systems using numerical data, genetic and evolutionary algorithms can be also used. As knowledge in fuzzy systems is structured in the form of rules we can also employ other methods for learning from numerical data, e. g. clustering algorithms with most often used fuzzy c-means algorithm. We can also utilize any method for generating rules from data, e. g. (Wang and Mendel 1992).

Having given the learning dataset of the pair $(\bar{\mathbf{x}}, d)$ where d is desired response of the system, we can use the following error measure:

$$Q(\bar{\mathbf{x}}, d) = \frac{1}{2} [\bar{y}(\bar{\mathbf{x}}) - d]^2. \quad (57)$$

Our goal is to minimize the error by determining the value of parameters $\bar{x}_i^r$, σ_i^r (antecedent Gaussian fuzzy sets) and $\bar{y}^r$ (consequent singleton). The parameters can be treated like weights in neural networks and in learning iteration (step) t the value of any weight can be defined by

$$p_i^k(t+1) = p_i^k(t) - \eta \frac{\partial Q(\bar{\mathbf{x}}, d; t)}{\partial p_i^k(t)}. \quad (58)$$

In the case of the simplest neuro-fuzzy system given by (37), we obtain the following formulas for gradients in recursive procedure (58)

$$\begin{aligned} \frac{\partial Q}{\partial \bar{x}_h^j} = & -2\varepsilon \frac{(\bar{x}_h^j - \bar{x}_h) \prod_{i=1}^n \mu_{A_i^j}(\bar{x}_i) \sum_{r=1}^N (\bar{y}^j - \bar{y}^r) \prod_{i=1}^n \mu_{A_i^r}(\bar{x}_i)}{\left\{ \sigma_h^j \sum_{r=1}^N \prod_{i=1}^n \mu_{A_i^r}(\bar{x}_i) \right\}^2}, \end{aligned} \quad (59)$$

$$\begin{aligned} \frac{\partial Q}{\partial \sigma_h^j} = & 2\varepsilon \frac{(\bar{x}_h^j - \bar{x}_h) \prod_{i=1}^n \mu_{A_i^j}(\bar{x}_i) \sum_{r=1}^N (\bar{y}^j - \bar{y}^r) \prod_{i=1}^n \mu_{A_i^r}(\bar{x}_i)}{\left\{ \sigma_h^j \sum_{r=1}^N \prod_{i=1}^n \mu_{A_i^r}(\bar{x}_i) \right\}^2}, \end{aligned} \quad (60)$$

$$\frac{\partial Q}{\partial \bar{y}^j} = \varepsilon \frac{\prod_{i=1}^n \mu_{A_i^j}(\bar{x}_i)}{\sum_{r=1}^N \prod_{i=1}^n \mu_{A_i^r}(\bar{x}_i)}, \quad (61)$$

where $\varepsilon = \bar{y} - d$ is an error calculated in the last layer of the system and $j = 1, \dots, n$, $h = 1, \dots, N$.

Criteria Isolines Method

In this section, we will first present the Akaike criterion, initially applied to the order estimation of autoregression processes, and next it will be adapted to evaluate the effectiveness of neuro-fuzzy systems. By the effectiveness of operation of a neuro-fuzzy system, we shall understand the precision (accuracy) of operation achieved by such a system, (expressed by mean squared error or by the number of wrongly classified samples) in the context of its size. By the system size we shall understand the number of all parameters that are subject to learning. We shall also present the concept of the so-called criteria isolines, which allow one to solve the problem of the compromise between the system accuracy and the number of parameters describing this system.

Neuro-fuzzy system should be characterized by the smallest possible error but at the same

time should be as simple as possible. It should be remembered that systems with smaller number of trained parameters are characterized among others by better generalization capabilities. Here, we should mention the so-called parsimony principle (Söderström and Stoica 1989). This principle is very useful when determining the appropriate order of the model. It may be formulated as follows: from between two alternative and satisfactory models, we shall choose the one which contains less independent parameters. This principle remains compliant with common sense: “Do not enter any additional parameters into the process description unless they are necessary”.

Estimation methods of the system order have been best developed for autoregression processes (Kay 1988; Marple Jr 1987). Time series $u(n)$, $u(n-1)$, $\dots$, $u(n-p)$ is an autoregression process of order p , if the difference equation is satisfied

$$u(n) + \alpha_1 u(n-1) + \dots + \alpha_p u(n-p) = e(n) \quad (62)$$

or equivalently

$$u(n) = - \sum_{k=1}^p \alpha_k u(n-k) + e(n), \quad (63)$$

where $\alpha_1, \dots, \alpha_p$ are process coefficient, while $e(n)$ is the white noise

$$E[e(n)e(m)] = \begin{cases} \sigma^2 & \text{dla } n = m \\ 0 & \text{dla } n \neq m \end{cases} \quad (64)$$

In the autoregression theory, criteria allowing one to estimate the order of predictor p , determining first the prediction error $\hat{Q}_p$ based on the learning sequence of the length M , are well known. The most important is the Akaike information criterion (AIC), Schwarz method and the final prediction error (FPE) method (Kay 1988).

The Akaike estimation order prediction method will be adapted now to the evaluation of fuzzy systems. Thanks to this, search for the desired fuzzy system based on two criteria (number of parameters and mean square error) will come down to one selected criterion, i. e.

AIC. It has been adapted for the needs of evaluation of neuro-fuzzy systems in the following form:

$$\text{AIC}(p, \hat{Q}_p) = M \ln \hat{Q}_p + 2p, \quad (65)$$

where p is the number of system parameters subject to learning (number of parameters of all membership functions and number of all weights if they occur in a given system), $\hat{Q}_p$ is the measure of error used in simulations, and M is the number of samples in a learning sequence. The product $M \cdot n$ may, therefore, be treated as a measure of size of the problem being solved. Table 2 contain the computed values of criteria for particular tested structures in case of the learning and testing sequence used in the polymerization problem (UCI n.d.). Figure 5 illustrates the coordinates of the points corresponding to particular neuro-fuzzy systems tested. The coordinate p defines the number of parameters of a given system; coordinate Q defines the error with which the system realized the problem to be solved.

The criteria isolines present constant values of the AIC criterion, with different values of the error and the number of parameters. Such an approach allows one to solve the problem of the compromise between the system operation error and the number of parameters describing this system. Points located on the criteria isolines with the same values of AIC criterion characterize the neuro-fuzzy systems making up the Pareto set. In the Pareto set, none of the two values of contradictory criteria may be improved (mean square error versus system size), without worsening the other one. Points located on the criteria isolines with the smallest values of AIC criterion characterize the neuro-fuzzy systems which have been called suboptimal ones. The suboptimal neuro-fuzzy systems presented in graphs ensure the smallest value of criteria within tested structures (the terminology “optimum systems” is not used as all possible structures have not been tested). From Table 2 it follows that for both learning and testing sequences the best, in the sense of the Akaike criterion, is the structure no. 1 (Larsen simplified) followed by the structure no. 29 (Zadeh with weights of rules).

Neuro-fuzzy Systems, Table 2 Value of the Akaike criterion for the polymerization problem for the learning and testing sequence

No.	Structure	Polymerization				
		p	Learning sequence		Testing sequence	
			Error	AIC	Error	AIC
1	Larsen simplified	42	0.0042	−299.09	0.0045	−294.26
2	Larsen simplified with weights of rules	48	0.0039	−292.27	0.0041	−288.77
3	Larsen simplified with weights of inputs and rules	66	0.0031	−272.34	0.0033	−267.97
4	Łukasiewicz simplified	42	0.0059	−275.30	0.0063	−270.70
5	Łukasiewicz simplified with weights of rules	48	0.0039	−292.27	0.0041	−288.77
6	Łukasiewicz simplified with weights of inputs and rules	66	0.0037	−259.96	0.0039	−256.27
7	Zadeh simplified	42	0.0049	−288.30	0.0053	−282.80
8	Zadeh simplified with weights of rules	48	0.0041	−288.77	0.0044	−283.83
9	Zadeh simplified with weights of inputs and rules	66	0.0038	−258.09	0.0040	−254.50
10	Binary	48	0.0063	−258.70	0.0067	−254.40
11	Binary with weights of rules	54	0.0054	−257.49	0.0057	−253.71
12	Binary with weights of inputs and rules	72	0.0036	−249.88	0.0038	−246.09
13	Larsen	48	0.0049	−276.30	0.0052	−272.14
14	Larsen with weights of rules	54	0.0043	−273.44	0.0045	−270.26
15	Larsen with weights of inputs and rules	72	0.0035	−251.85	0.0038	−246.09
16	Łukasiewicz	48	0.0065	−256.52	0.0069	−252.34
17	Łukasiewicz with weights of rules	54	0.0041	−276.77	0.0045	−270.26
18	Łukasiewicz with weights of inputs and rules	72	0.0038	−246.09	0.0041	−240.77
19	Mamdani	48	0.0041	−288.77	0.0043	−285.44
20	Mamdani with weights of rules	54	0.0039	−280.27	0.0042	−275.09
21	Mamdani with weights of inputs and rules	72	0.0034	−253.88	0.0037	−247.96
22	Reichenbach	48	0.0040	−290.50	0.0044	−283.83
23	Reichenbach with weights of rules	54	0.0037	−283.96	0.0039	−280.27
24	Reichenbach with weights of inputs and rules	72	0.0034	−253.88	0.0037	−247.96
25	Willmott	48	0.0056	−266.95	0.0060	−262.12
26	Willmott with weights of rules	54	0.0047	−267.21	0.0049	−264.30
27	Willmott with weights of inputs and rules	72	0.0039	−244.27	0.0043	−237.44
28	Zadeh	48	0.0038	−294.09	0.0043	−285.44
29	Zadeh with weights of rules	54	0.0030	−298.64	0.0033	−291.97
30	Zadeh with weights of inputs and rules	72	0.0028	−267.47	0.0031	−260.34
31	Takagi–Sugeno	60	0.0034	−277.88	0.0036	−273.88
32	Takagi–Sugeno with weights of rules	66	0.0031	−272.34	0.0034	−265.88
33	Takagi–Sugeno with weights of inputs and rules	84	0.0030	−238.64	0.0033	−231.97

Future Directions

Mankind creates and collects large volumes of data nowadays. We deal with an enormous amount of, often multidimensional, data. We can find such data in medicine and health care, business, manufacturing and finance, science and telecommunication. Neuro-fuzzy systems have to

cope with these large, often multidimensional and incomplete, data sets.

In case of real world data, learning systems have to face with a problem of missing data, i. e. missing certain features (input variables). Such shortages can be replaced by the average or modal value, by a value provided by the user or by a value resulting from other records. Good solution in this case is provided by information

- Bubnicki Z (2002b) A unified approach to descriptive and prescriptive concepts in uncertain decision systems. *Syst Anal Model Simul* 42(3):331–342. Gordon and Breach Science Publishers, Newark
- Chen MY, Linkens DA (2001) A systematic neuro-fuzzy modeling framework with application to material property prediction. *IEEE Trans Fuzzy Syst* 31:781–790
- Chuang CC, Su SF, Chen SS (2001) Robust TSK fuzzy modeling for function approximation with outliers. *IEEE Trans Fuzzy Syst* 9:810–821
- Corcoran AL, Sen S (1994) Using real-valued genetic algorithms to evolve rule sets for classification. In: *Proceedings of the first IEEE conference on evolutionary computation*. Orlando, June, pp 120–124
- Czogała E, Łeński J (2000) *Fuzzy and neuro-fuzzy intelligent systems*. Physica-Verlag Company, Heidelberg/New York
- Duan JC, Chung FL (2001) Cascaded fuzzy neural network based on syllogistic fuzzy reasoning. *IEEE Trans Fuzzy Syst* 9:293–306
- Eubank RL (1999) *Nonparametric regression and spline smoothing*. Marcel Dekker, New York
- Fodor JC (1991) On fuzzy implication operators. *Fuzzy Sets Syst* 42(3):293–300. Elsevier, Amsterdam
- Fogel DB (1995) *Evolutionary computation: towards a new philosophy of machine intelligence*. IEEE Press, New York
- Fuller R (2000) *Introduction to neuro-fuzzy systems, advances in soft computing*. Physica-Verlag, New York
- Gaweda AE, Żurada JM (2000) Fuzzy neural network with relational fuzzy rules. In: *Proceedings of the international joint conference on neural networks IJCNN'2000*, vol 5, pp 3–8, Como, Italy, July 23–27
- Gaweda AE, Żurada JM (2001) Data-driven design of fuzzy system with relational input partition. In: *Proceedings of the international conference on fuzzy systems FUZZ-IEEE'2001*, Melbourne, Australia, December 2–5
- Hirota K (1993) *Industrial applications of fuzzy technology*. Springer, Tokyo/Berlin/Heidelberg/New York
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23: 665–685
- Jang JSR, Sun CT (1995) Neuro-fuzzy modeling and control. *Proc IEEE* 83:378–406
- Jang JS, Sun CT, Mizutani E (1997) *Neuro-fuzzy and soft computing*. Prentice Hall, Englewood Cliffs
- Juang C-F, Lin C-T (1998) An on-line self-constructing neural fuzzy inference network and its applications. *IEEE Trans Fuzzy Syst* 6:12–32
- Kacprzyk J (1997) *Multistage fuzzy control*. Wiley, Chichester
- Kasabov N (1996) *Foundations of neural networks, fuzzy systems and knowledge engineering*. The MIT Press CA, Cambridge
- Kasabov N (2002) DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE Trans Fuzzy Syst* 10:144–154
- Kay SM (1988) *Modern spectral estimation. Theory and application*. Prentice Hall, Englewood Cliffs
- Kecman V (2001) *Learning and soft computing*. MIT Press, Cambridge
- Klement EP, Mesiar R, Pap E (2000) *Triangular norms*. Kluwer, Dordrecht
- Lee K-M, Kwak D-H, Lee-Kwang H (1994) A fuzzy neural network model for fuzzy inference and rule tuning. *Int J Uncertainty Fuzziness Knowledge-Based Syst* 2(3): 265–277
- Lee K-M, Kwak D-H, Lee-Kwang H (1996) Fuzzy inference neural network for fuzzy model tuning. *IEEE Trans Syst Man Cybern Part B* 26(4):637–645
- Lin CT (1994) *Neural fuzzy control systems with structure and parameter learning*. World Scientific, Singapore
- Lin Y, Cunningham GA III (1995) A new approach to fuzzy-neural system modeling. *IEEE Trans Fuzzy Syst* 3:190–198
- Lin CT, Lee CSG (1991) Neural-network-based fuzzy logic control and decision system. *IEEE Trans Comput* 40:1320–1336
- Lin CT, Lee GCS (1997) *Neural fuzzy systems a neuro-fuzzy synergism to intelligent systems*. Prentice Hall, Englewood Cliffs
- Lin CT, Lu YC (1995) A neural fuzzy systems with linguistic teaching signals. *IEEE Trans Fuzzy Syst* 3: 169–189
- Marple SL Jr (1987) *Digital spectral analysis with applications*. Prentice Hall, Englewood Cliffs
- Michalewicz Z (1992) *Genetic algorithms + data structures = evolution programs*. Springer, Berlin
- Mouzouris GC, Mendel JM (1997) Nonsingleton fuzzy logic systems: theory and application. *IEEE Trans Fuzzy Syst* 5(1):56–71
- Nauck D, Kruse R (1996) Designing neuro-fuzzy systems through back-propagation. In: *Pedrycz W (ed) Fuzzy modeling: paradigms and practice*. Kluwer, Boston, pp 203–228
- Nauck D, Kruse R (1999) Neuro-fuzzy systems for function approximation. *Fuzzy Sets Syst* 101:261–271
- Nauck D, Klawon F, Kruse R (1997) *Foundations of neuro-fuzzy systems*. Wiley, Chichester
- Nie J, Linkens D (1995) *Fuzzy-neural control. principles, algorithms and applications*. Prentice Hall, New York/London
- Pagan A, Ullah A (1999) *Nonparametric econometrics*. Cambridge University Press, London
- Pawlak Z (1982) Rough sets. *Int J Inform Comput Sci* 11(341)
- Pawlak Z (1991) *Rough sets. Theoretical aspects of reasoning about data*. Kluwer, Dordrecht
- Pedrycz W (1992) Fuzzy neural networks with reference neurons as pattern classifiers. *IEEE Trans Neural Netw* 3(5):770–775
- Roubos H, Setnes M (2001) Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Trans Fuzzy Syst* 9:516–524
- Rutkowska D (2002) *Neuro-fuzzy architectures and hybrid learning*. Springer, Heidelberg
- Rutkowski L (2004) *Flexible neuro-fuzzy systems*. Kluwer, Norwell
- Rutkowski L, Cpałka K (2000) *Flexible structures of neuro-fuzzy systems, quo Vadis computational*

- intelligence, studies in fuzziness and soft computing, vol 54. Springer, Berlin, pp 479–484
- Rutkowski L, Cpalka K (2003) Flexible neuro-fuzzy systems. *IEEE Trans Neural Netw* 14:554–574
- Rutkowski L, Rafajłowicz E (1989) On global rate of convergence of some nonparametric identification procedures. *IEEE Trans Autom Control* AC-34(10): 1089–1091
- Söderström T, Stoica P (1989) System identification. Prentice-Hall, London
- Tadeusiewicz R (1993) Neural networks RM. Academic Publishing House, Warsaw (in Polish)
- Tadeusiewicz R (1998) Elementary introduction to neural networks with computer programs. Academic Publishing House, Warsaw (in Polish)
- Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. *IEEE Trans Syst Man Cybern* 15:116–132
- UCI. Respository of machine learning databases. Available online: <http://ftp.ics.uci.edu/pub/machine-learning-databases/>
- Wang LX (1994) Adaptive fuzzy systems and control. PTR Prentice Hall, Englewood Cliffs
- Wang JS, Lee CSG (2002) Self-adaptive neuro-fuzzy inference systems for classification applications. *IEEE Trans Fuzzy Syst* 10:790–802
- Wang LX, Mendel JM (1992) Generating fuzzy rules by learning from examples. *IEEE Trans Syst Man Cybern* 22(6):1414–1427
- Yager RR (1990) Fuzzy logic controller structures. *Proc SPIE Symp Laser Sci Optics Appl*:368–378
- Yager RR (1992) A general approach to rule aggregation in fuzzy logic control. *Appl Intell* 2:333–351
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Žurada JM (1992) Introduction to artificial neural systems. West Publishing Company

Possibility Theory

Didier Dubois and Henri Prade
IRIT-CNRS, Universite Paul Sabatier, Toulouse,
France

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Historical Background
Basic Notions of Possibility Theory
Qualitative Possibility Theory
Quantitative Possibility Theory
Probability-Possibility Transformations
Applications and Future Directions
Bibliography

Glossary

Guaranteed Possibility A guaranteed possibility measure is a set function (decreasing in the wide sense) that returns the minimum of a possibility distribution over a subset representing an event. While possibility measures evaluate the consistency of the information between an event and the available information represented by the underlying possibility distribution, guaranteed possibility measures capture another view of the idea of possibility related to the idea of (guaranteed) feasibility or sufficiency condition.

Necessity Measure A necessity measure is a set function, associated by duality to a possibility measure through a relation expressing that an event is all the more necessarily true (all the more certain) as the opposite event is less possible. A necessity measure estimates to what extent the information represented by the underlying possibility distribution entails the occurrence of the event.

Possibilistic Logic Standard possibilistic logic is a weighted logic where formulas are pairs made of a classical logical formula and a weight that acts as a lower bound of the necessity of the logical formula. Extended possibilistic logics may include formulas weighted in terms of lower bounds of possibility or guaranteed possibility measures.

Possibility Distribution A possibility distribution restricts a set of possible values for a variable of interest in an elastic way. It is represented by a mapping from a universe gathering the potential values of the variable to a scale such as the unit interval of the real line or a finite linearly ordered set, expressing to what extent each value is possible for the variable. Thus, a possibility distribution restricts a set of more or less possible values belonging to a universe that may be also ordered such as a subpart of real line for a numerical variable, or not ordered if, for instance, the variable takes its value in the set of interpretations of a logical language. This may be used for representing uncertainty if the restriction pertains to possible values for an ill-known state of the world, or for representing preferences if the restriction encodes a set of values that are considered as more or less satisfactory for some purpose.

Possibility Measure A possibility measure is a set function (increasing in the wide sense) that returns the maximum of a possibility distribution over a subset representing an event.

Definition of the Subject and Its Importance

Possibility theory is the simplest uncertainty theory devoted to the modeling of incomplete information. It is characterized by the use of two basic dual set functions that respectively grade the possibility and the necessity of events. Possibility

theory lies at the crossroads between fuzzy sets, probability, and nonmonotonic reasoning. Possibility theory is closely related to fuzzy sets if one considers that a possibility distribution is a particular fuzzy set (of mutually exclusive) possible values. However, fuzzy sets and fuzzy logic are primarily motivated by the representation of gradual properties while possibility theory handles the uncertainty of classical (or fuzzy) propositions. Possibility theory can be cast either in an ordinal or in a numerical setting. Qualitative possibility theory is closely related to belief revision theory and commonsense reasoning with exception-tainted knowledge in artificial intelligence. It has been axiomatically justified in a decision-theoretic framework in the style of Savage, thus providing a foundation for qualitative decision theory. Quantitative possibility theory is the simplest framework for statistical reasoning with imprecise probabilities. As such, it has close connections with random set theory and confidence intervals and can provide a tool for uncertainty propagation with limited statistical or subjective information.

Introduction

Possibility theory is an uncertainty theory devoted to the handling of incomplete information. To a large extent, it is similar to probability theory because it is based on set functions. It differs from the latter by the use of a pair of dual set functions (possibility and necessity measures) instead of only one. Besides, it is not additive and makes sense on ordinal structures. The name “Theory of Possibility” was coined by Zadeh (1978), who was inspired by a paper by Gaines and Kohout (1975). In Zadeh’s view, possibility distributions were meant to provide graded semantics to natural language statements. However, possibility and necessity measures can also be the basis of a full-fledged representation of partial belief that parallels probability. It can be seen either as a coarse, non-numerical version of probability theory or a framework for reasoning with extreme probabilities or yet a simple

approach to reasoning with imprecise probabilities (Dubois and Prade 1998).

After reviewing pioneering contributions to possibility theory, we recall its basic concepts and present the two main directions along which it has developed: the qualitative and quantitative settings. Both approaches share the same basic “maxitivity” axiom. They differ when it comes to conditioning and to independence notions.

Historical Background

Zadeh was not the first scientist to speak about formalizing notions of possibility. The modalities *possible* and *necessary* have been used in philosophy at least since the Middle-Ages in Europe, based on Aristotle’s works. More recently, they became the building blocks of modal logics that emerged at the beginning of the twentieth century from the works of C.I. Lewis (see Hughes and Cresswell 1968). In this approach, possibility and necessity are all-or-nothing notions and handled at the syntactic level. More recently, and independently from Zadeh’s view, the notion of possibility, as opposed to probability, was central in the works of one economist, Shackle, and in those of two philosophers, D. Lewis and L. J. Cohen.

G. L. S. Shackle

A graded notion of possibility was introduced as a full-fledged approach to uncertainty and decision in the 1940s–1970s by the English economist G. L. S. Shackle (Shackle 1961), who called *degree of potential surprise* of an event its degree of impossibility, that is, the degree of necessity of the opposite event. Shackle’s notion of possibility is basically epistemic; it is a “character of the chooser’s particular state of knowledge in his present.” Impossibility is understood as disbelief. Potential surprise is valued on a disbelief scale, namely, a positive interval of the form $[0, y^*]$, where y^* denotes the absolute rejection of the event to which it is assigned. In case everything is possible, all mutually exclusive hypotheses have zero surprise. At least one elementary hypothesis must carry zero potential surprise.

The degree of surprise of an event, a set of elementary hypotheses, is the degree of surprise of its least surprising realization. The disbelief notion introduced later by Spohn (Spohn 1990) employs the same type of convention as potential surprise but using the set of natural integers as a disbelief scale. Shackle also introduces a notion of conditional possibility, whereby the degree of surprise of a conjunction of two events A and B is equal to the maximum of the degree of surprise of A and of the degree of surprise of B , should A prove true.

D. Lewis

In his 1973 book (Lewis 1973), the philosopher David Lewis considers a graded notion of possibility in the form of a relation between possible worlds he calls *comparative possibility*. He equates this concept of possibility to a notion of similarity between possible worlds. This non-symmetric notion of similarity is also comparative and is meant to express statements of the form *a world j is at least as similar to world i as world k is*. Comparative similarity of j and k with respect to i is interpreted as the comparative possibility of j with respect to k viewed from world i . Such relations are assumed to be complete preorderings and are instrumental in defining the truth conditions of counterfactual statements. Comparative possibility relations $\geq_{\Pi}$ obey the key axiom: for all events A, B, C ,

$$A \geq_{\Pi} B \text{ implies } C \cup A \geq_{\Pi} C \cup B.$$

This axiom was later independently proposed by the first author (Dubois 1986) in an attempt to derive a possibilistic counterpart to comparative probabilities. The connection between numerical possibility and similarity has been investigated by Sudkamp (2002).

L. J. Cohen

A framework very similar to the one of Shackle was proposed by the philosopher L. J. Cohen (1977), who considered the problem of legal reasoning. He introduced so-called Baconian probabilities understood as degrees of provability. The idea is that it is hard to prove someone guilty at the court of law by means of pure statistical

arguments. The basic feature of degrees of provability is that a hypothesis and its negation cannot both be provable together to any extent (the contrary being a case for inconsistency). Such degrees of provability coincide with necessity measures.

L. A. Zadeh

In his seminal paper (Zadeh 1978), Zadeh proposed an interpretation of membership functions of fuzzy sets as possibility distributions encoding flexible constraints induced by natural language statements. Zadeh articulated the relationship between possibility and probability, noticing that what is probable must preliminarily be possible. However, the view of possibility degrees developed in his paper refers to the idea of graded feasibility (degrees of ease, as in the example of “how many eggs can Hans eat for his breakfast”) rather than to the epistemic notion of plausibility laid bare by Shackle. Nevertheless, the key axiom of “maxitivity” for possibility measures is highlighted. In two subsequent articles (Zadeh 1979, 1982), Zadeh acknowledged the connection between possibility theory, belief functions, and upper/lower probabilities and proposed their extensions to fuzzy events and fuzzy information granules.

Basic Notions of Possibility Theory

The basic building blocks of possibility theory were first extensively described in the authors’ books (Dubois and Prade 1980, 1988) (see also Klir and Folger 1988). Let S be a set of states of affairs (or descriptions thereof) or states for short. A possibility distribution is a mapping π from S to a totally ordered scale L , with top 1 and bottom 0, such as the unit interval. The function π represents the state of knowledge of an agent (about the actual state of affairs) distinguishing what is plausible from what is less plausible, what is the normal course of things from what is not, what is surprising from what is expected. It represents a flexible restriction on what is the actual state with the following conventions (similar to probability but opposite to Shackle’s potential surprise scale):

- $\pi(s) = 0$ means that state s is rejected as impossible.
- $\pi(s) = 1$ means that state s is totally possible (= plausible).

If S is exhaustive, at least one of the elements of S should be the actual world, so that $\exists s, \pi(s) = 1$ (normalization). Distinct values may simultaneously have a degree of possibility equal to 1.

Possibility theory is driven by the principle of minimal specificity. It states that any hypothesis not known to be impossible cannot be ruled out. A possibility distribution π is said to be at least as specific as another π' if and only if for each state of affairs s : $\pi(s) \leq \pi'(s)$ (Yager 1983). Then, π is at least as restrictive and informative as π' .

In the possibilistic framework, extreme forms of partial knowledge can be captured, namely,

- Complete knowledge: for some s_0 , $\pi(s_0) = 1$ and $\pi(s) = 0$, $\forall s \neq s_0$ (only s_0 is possible)
- Complete ignorance: $\pi(s) = 1$, $\forall s \in S$, (all states are possible)

Given a simple query of the form “does event A occur?” where A is a subset of states, the response to the query can be obtained by computing degrees of possibility and necessity, respectively (if the possibility scale $L = [0,1]$):

$$\Pi(A) = \sup_{s \in A} \pi(s); N(A) = \inf_{s \notin A} 1 - \pi(s).$$

$\Pi(A)$ evaluates to what extent A is consistent with π , while $N(A)$ evaluates to what extent A is certainly implied by π . The possibility-necessity duality is expressed by $N(A) = 1 - \Pi(A^c)$, where A^c is the complement of A . Generally, $\Pi(S) = N(S) = 1$ and $\Pi(\emptyset) = N(\emptyset) = 0$. Possibility measures satisfy the basic “maxitivity” property $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$. Necessity measures satisfy an axiom dual to that of possibility measures, namely, $N(A \cap B) = \min(N(A), N(B))$. On infinite spaces, these axioms must hold for infinite families of sets. Besides, a modal logic counterpart of possibility theory, called MEL (Banerjee and Dubois 2009, 2014),

with semantics in terms of epistemic states has been proposed for binary possibility and necessity.

Human knowledge is often expressed in a declarative way using statements to which belief degrees are attached. It corresponds to expressing constraints the world is supposed to comply with. Certainty-qualified pieces of uncertain information of the form “ A is certain to degree α ” can then be modeled by the constraint $N(A) \geq \alpha$. The least specific possibility distribution reflecting this information is (Dubois and Prade 1988)

$$\pi_{(A,\alpha)}(s) = \begin{cases} 1, & \text{if } s \in A \\ 1 - \alpha & \text{otherwise} \end{cases} \quad (1)$$

Acquiring further pieces of knowledge leads to updating $\pi_{(A,\alpha)}$ into some $\pi < \pi_{(A,\alpha)}$.

Apart from Π and N , a measure of *guaranteed possibility* or *sufficiency* can be defined (Dubois et al. 2000b): $\Delta(A) = \inf_{s \in A} \pi(s)$. It estimates to what extent *all* states in A are actually possible according to evidence. In contrast, Π appears to be a measure of *potential* possibility. $\Delta(A)$ can be used as a degree of evidential support for A . Uncertain statements of the form “ A is possible to degree β ” often mean that all realizations of A are possible to degree β . They can then be modeled by the constraint $\Delta(A) \geq \beta$. It corresponds to the idea of observed evidence. This type of information is better exploited by assuming an informational principle opposite to the one of minimal specificity, namely, any situation not yet observed is tentatively considered as impossible. This is similar to closed-world assumption. The most specific distribution $\delta_{(A,\beta)}$ in agreement with $\Delta(A) \geq \beta$ is

$$\delta_{(A,\beta)}(s) = \begin{cases} \beta, & \text{if } s \in A \\ 0 & \text{otherwise.} \end{cases}$$

Acquiring further pieces of evidence leads to updating $\delta_{(A,\beta)}$ into some wider distribution $\delta > \delta_{(A,\beta)}$. Such evidential support functions do not behave with the same conventions as possibility distributions: $\delta(s) = 1$ means that S is guaranteed to be possible, because of a high evidential support,

while $\delta(s) = 0$ only means that S has not been observed yet (hence is of unknown possibility). Distributions δ are generally not normalized to 1 and serve as lower bounds to possibility distributions π (because what is observed must be possible). Such a bipolar representation of information using pairs (δ, π) may provide a natural interpretation of interval-valued fuzzy sets. Note that possibility distributions induced from certainty-qualified pieces of knowledge combine conjunctively, by discarding possible states, while evidential support distributions induced by possibility-qualified pieces of evidence combine disjunctively, by accumulating possible states.

Notions of conditioning and independence were studied for possibility measures. Conditional possibility is defined similarly to probability theory using a Bayesian-like equation of the form (Dubois and Prade 1988)

$$\Pi(B \cap A) = \Pi(B \mid A) * \Pi(A).$$

However, in the ordinal setting, the operation $*$ cannot be a product and is changed into the minimum. In the numerical setting, there are several ways to define conditioning, not all of which have this form (Walley 1996). There are several variants of possibilistic independence (De Cooman 1997; De Campos and Huete 1999; Dubois et al. 1997). Generally, independence in possibility theory is neither symmetric nor insensitive to negation. For non-Boolean variables, independence between events is not equivalent to independence between variables. Joint possibility distributions on Cartesian products of domains can be represented by means of graphical structures similar to Bayesian networks for joint probabilities (see Borgelt et al. 2000b; Benferhat et al. 2002c). Such graphical structures can be taken advantage of for evidence propagation (Ben Amor et al. 2003) or learning (Borgelt and Kruse 2003).

Qualitative Possibility Theory

This section is restricted to the case of a finite state space S , supposed to be the set of interpretations of a formal propositional language. In other

words, S is the universe induced by Boolean attributes. A plausibility ordering is a complete pre-order of states denoted by $\geq_\pi$ which induces a well-ordered partition $\{E_1, \dots, E_n\}$ of S . It is the comparative counterpart of a possibility distribution π , i.e., $s \geq_\pi s'$ if and only if $\pi(s) \geq \pi(s')$. Indeed it is more natural to expect that an agent will supply ordinal rather than numerical information about his beliefs. By convention, E_1 contains the most normal states of fact, E_n the least plausible or most surprising ones. Denoting by $\max(A)$ any most plausible state $s_0 \in A$, ordinal counterparts of possibility and necessity measures (Dubois 1986) are then defined as follows: $\{s\} \geq \emptyset$, for all $s \in S$ and

$$A \geq_\Pi B \text{ if and only if } \max(A) \geq_\pi \max(B)$$

$$A \geq_N B \text{ if and only if } \max(B^c) \geq_\pi \max(A^c).$$

Possibility relations $\geq_\pi$ are those of Lewis and satisfy the characteristic property

$$A \geq_\Pi B \text{ implies } C \cup A \geq_\Pi C \cup B$$

while necessity relations can also be defined as $A \geq_N B$ if and only if $B^c \geq_\Pi A^c$ and satisfy a similar axiom:

$$A \geq_N B \text{ implies } C \cap A \geq_N C \cap B.$$

The latter coincide with epistemic entrenchment relations in the sense of belief revision theory (Gärdenfors 1998; Dubois and Prade 1991a). Conditioning a possibility relation $\geq_\Pi$ by a non-impossible event $C \geq_\Pi \phi$ means deriving a relation $\geq_{\Pi}^C$ such that

$$A \geq_{\Pi}^C B \text{ if and only if } A \cap C \geq_\Pi B \cap C.$$

The notion of independence for comparative possibility theory was studied in Dubois et al. (1997), for independence between events, and Ben Amor et al. (2002) between variables.

Nonmonotonic Inference

Suppose S is equipped with a plausibility ordering. The main idea behind qualitative possibility

theory is that the state of the world is always believed to be as normal as possible, neglecting less normal states. $A \geq_{\Pi} B$ really means that there is a normal state where A holds that is at least as normal as any normal state where B holds. The dual case $A \geq_N B$ is intuitively understood as “ A is at least as certain as B ,” in the sense that there are states where B fails to hold that are at least as normal as the most normal state where A does not hold. In particular, the events accepted as true are those which are true in all the most plausible states, namely, the ones such that $A >_N \phi$. These assumptions lead us to interpret the plausible inference $A \approx B$ of a proposition B from another A , under a state of knowledge $\geq_{\Pi}$ as follows: *B should be true in all the most normal states where A is true*, which means $B >_{\Pi}^A B^c$ in terms of ordinal conditioning, that is, $A \cap B$ is more plausible than $A \cap B^c$. $A \approx B$ also means that the agent considers B as an accepted belief in the context A .

This kind of inference is nonmonotonic in the sense that $A \approx B$ does not always imply $A \cap C \approx B$ for any additional information C . This is similar to the fact that a conditional probability $P(B | A \cap C)$ may be low even if $P(B | A)$ is high. The properties of the consequence relation $| \approx$ are now well understood and are precisely the ones laid bare by Lehmann and Magidor (1992) for their so-called rational inference. Monotonicity is only partially restored: $A \approx B$ implies $A \cap C \approx B$ holds provided that $A \approx C^c$ does not hold (i.e., that states where A is true do not typically violate C). This property is called *rational monotony* and, along with some more standard ones (like closure under conjunction), characterizes default possibilistic inference $| \approx$. In fact, the set $\{B, A \approx B\}$ of accepted beliefs in the context A is deductively closed, which corresponds to the idea that the agent reasons with accepted beliefs in each context as if they were true, until some event occurs that modifies this context. This closure property is enough to justify a possibilistic approach (Dubois et al. 2004a), and adding the rational monotonicity property ensures the existence of a single possibility relation generating the consequence relation $| \approx$ (Benferhat et al. 1997).

Rather than being constructed from scratch, plausibility orderings can be generated by a set of if-then rules tainted with unspecified exceptions. This set forms a knowledge base supplied by an agent. Each rule “if A then B ” is understood as a constraint of the form $A \cap B >_{\Pi} A \cap B^c$ on possibility relations. There exists a single minimally specific element in the set of possibility relations satisfying all constraints induced by rules (unless the latter are inconsistent). It corresponds to the most compact plausibility ranking of states induced by the rules (Benferhat et al. 1997). This ranking can be computed by an algorithm originally proposed by Pearl (1990).

Possibilistic Logic

Qualitative possibility relations can be represented by (and only by) possibility measures ranging on any totally ordered set L (especially a finite one) (Dubois 1986). This absolute representation on an ordinal scale is slightly more expressive than the purely relational one. When the finite set S is large and generated by a propositional language, qualitative possibility distributions can be efficiently encoded in possibilistic logic (Dubois et al. 1994). A possibilistic logic base K is a set of pairs (ϕ, α) , where ϕ is a Boolean expression and α is an element of L . This pair encodes the constraint $N(\phi) \geq \alpha$ where $N(\phi)$ is the degree of necessity of the set of models of ϕ . Each prioritized formula (ϕ, α) has a fuzzy set of models (described by formula (1) where A is replaced by the set of models of ϕ), and the fuzzy intersection of the fuzzy sets of models of all prioritized formulas in K yields the associated plausibility ordering on S , K being viewed as the conjunction of its possibilistic logic formulas.

Syntactic deduction from a set of prioritized clauses is achieved by refutation using an extension of the standard resolution rule, whereby $(\phi \vee \psi, \min(\alpha, \beta))$ can be derived from $(\phi \vee \xi, \alpha)$ and $(\psi \vee \neg \xi, \beta)$. This rule, which evaluates the validity of an inferred proposition by the validity of the weakest premise, goes back to Theophrastus, a disciple of Aristotle. Possibilistic logic is an inconsistency-tolerant extension of propositional logic that provides a natural semantic setting for

mechanizing nonmonotonic reasoning (Benferhat et al. 1998), with a computational complexity close to that of propositional logic. See Dubois and Prade (2004, 2015) for a detailed introduction to possibilistic logic, its syntactic and semantic aspects, different extensions that involve time, for instance, or that encode constraints on lower bounds of possibility or guaranteed possibility measures, or that perform multiple source information fusion. See Dubois and Prade (2007) for a sketch of further extensions for handling groups of agents' beliefs and mutual beliefs. Generalized possibilistic logic allows not only for the conjunction but also for the disjunction and the negation of formulas (ϕ, α) ; it has a semantics in terms of *sets* of possibility distributions and provides a clear semantics for ASP logic programming formulas (Dubois et al. 2012). Let us also mention a recent noticeable result that establishes that any monotonously increasing set function can be characterized on a finite universe by a *set* of possibility measures (Dubois et al. 2013).

Another compact representation of qualitative possibility distributions is the possibilistic directed graph, which uses the same conventions as Bayesian nets but relies on an ordinal notion of conditional possibility (Dubois and Prade 1988)

$$\Pi(B | A) = \begin{cases} 1, & \text{if } \Pi(B \cap A) = \Pi(A) \\ \Pi(B \cap A) & \text{otherwise.} \end{cases}$$

Joint possibility distributions can be decomposed into a conjunction of conditional possibility distributions (using minimum) in a way similar to Bayes nets (Benferhat et al. 2002a). It is based on a symmetric notion of qualitative independence $\Pi(B \cap A) = \min(\Pi(A), \Pi(B))$ that is weaker than the causal-like condition $\Pi(B | A) = \Pi(B)$ (Dubois et al. 1997). The reference (Ben Amor and Benferhat 2005) investigates the properties of qualitative independence that enable local inferences to be performed in possibilistic nets.

Decision-Theoretic Foundations

Zadeh (1978) hinted that “since our intuition concerning the behavior of possibilities is not very reliable,” our understanding of them

“would be enhanced by the development of an axiomatic approach to the definition of subjective possibilities in the spirit of axiomatic approaches to the definition of subjective probabilities.” Decision-theoretic justifications of qualitative possibility were recently devised, in the style of Savage (1972). On top of the set of states, assume there is a set X of consequences of decisions. A decision, or act, is modeled as a mapping f from S to X assigning to each state s its consequence $f(s)$. The axiomatic approach consists in proposing properties of a preference relation $\succsim$ between acts so that a representation of this relation by means of a preference functional $W(f)$ is ensured, that is, act f is as good as act g (denoted $f \succsim g$) if and only if $W(f) \geq W(g)$. $W(f)$ depends on the agent's knowledge about the state of affairs, here supposed to be a possibility distribution π on S , and the agent's goal, modeled by a utility function u on X . Both the utility function and the possibility distribution map to the same finite chain L . A pessimistic criterion $W_{\pi}^{-}(f)$ is of the form

$$W_{\pi}^{-}(f) = \min_{s \in S} \max(n(\pi(s)), u(f(s)))$$

where n is the order-reversing map of L , $n(\pi(s))$ the degree of certainty that the state is not s (hence the degree of surprise of observing s), $u(f(s))$ the utility of choosing act f in state s . $W_{\pi}^{-}(f)$ is all the higher as all states are either very surprising or have high utility. This criterion is actually a prioritized extension of the Wald maximin criterion. The latter is recovered if $\pi(s) = 1$ (top of L) $\forall s \in S$. According to the pessimistic criterion, acts are chosen according to their worst consequences, restricted to the most plausible states $S^* = \{s, \pi(s) \geq n(W_{\pi}^{-}(f))\}$. The optimistic counterpart of this criterion is

$$W_{\pi}^{+}(f) = \max_{s \in S} \min(\pi(s), u(f(s))).$$

$W_{\pi}^{+}(f)$ is all the higher as there is a very plausible state with high utility. The optimistic criterion was first proposed by Yager (Yager 1979) and the pessimistic criterion by Whalen

(Whalen 1984). These optimistic and pessimistic possibilistic criteria are particular cases of a more general criterion based on the Sugeno integral (Grabisch et al. 2000) specialized to possibility and necessity of fuzzy events (Zadeh 1978; Dubois and Prade 1980):

$$S_{\gamma,u}(f) = \max_{\lambda \in L} \min(\lambda, \gamma(F_\lambda))$$

where $F_\lambda = \{s \in S, u(f(s)) \geq \lambda\}$, γ is a monotonic set function that reflects the decision-maker attitude in front of uncertainty, $\gamma(A)$ is the degree of confidence in event A . If $\gamma = \Pi$, then $S_{\Pi,u}(f) = W_\pi^+(f)$. Similarly, if $\gamma = N$, then $S_{N,u}(f) = W_\pi^-(f)$.

For any acts f, g and any event A , let fAg denote an act consisting of choosing f if A occurs and g if its complement occurs. Let $f \wedge g$ (resp. $f \vee g$) be the act whose results yield the worst (resp. best) consequence of the two acts in each state. Constant acts are those whose consequence is fixed regardless of the state. A result in (Dubois et al. 2000c, 2001) provides an act-driven axiomatization of these criteria and enforces possibility theory as a “rational” representation of uncertainty for a finite state space S :

Theorem 1 *Suppose the preference relation $\succsim$ on acts obeys the following properties:*

1. $(X^S, \succsim)$ is a complete preorder.
2. There are two acts such that $f \succ g$.
3. $\forall A, \forall g$, and h constant, $\forall f, g \succsim h$ implies $gAf \succsim hAf$.
4. If f is constant, $f \succ h$ and $g \succ h$ imply $f \wedge g \succ h$.
5. If f is constant, $h \succ f$ and $h \succ g$ imply $h \succ f \vee g$.

Then there exists a finite chain L , an L -valued monotonic set function γ on S , and an L -valued utility function u , such that $\succsim$ is representable by a Sugeno integral of $u(f)$ with respect to γ . Moreover, γ is a necessity (resp. possibility) measure as soon as property (iv) (resp. (v)) holds for all acts. The preference functional is then $W_\pi^-(f)$ (resp. $W_\pi^+(f)$).

Axioms (4–5) contradict expected utility theory. They become reasonable if the value scale is finite, decisions are one shot (no compensation), and provided that there is a big step between any level in the qualitative value scale and the adjacent ones. In other words, the preference pattern $f \succ h$ always means that f is significantly preferred to h , to the point of considering the value of h negligible in front of the value of f . The above result provides decision-theoretic foundations of possibility theory, whose axioms can thus be tested from observing the choice behavior of agents. See Dubois et al. (2003a) for another approach to comparative possibility relations, more closely relying on Savage axioms but giving up any comparability between utility and plausibility levels. The drawback of these and other qualitative decision criteria is their lack of discrimination power (Dubois and Fargier 2003). To overcome it, refinements of possibilistic criteria were proposed, based on lexicographic schemes (Fargier and Sabbadin 2005). These new criteria turn out to be representable by a classical (but big-stepped) expected utility criterion.

Quantitative Possibility Theory

The phrase “quantitative possibility” refers to the case when possibility degrees range in the unit interval. In that case, a precise articulation between possibility and probability theories is useful to provide an interpretation to possibility and necessity degrees. Several such interpretations can be consistently devised. See Dubois (2006) for a detailed survey. A degree of possibility can be viewed as an upper probability bound (Dubois and Prade 1992), and a possibility distribution can be viewed as a likelihood function (Dubois et al. 1997b). A possibility measure is also a special case of a Shafer plausibility function (Shafer 1987). Following a very different approach, possibility theory can account for probability distributions with extreme values, infinitesimal (Spohn 1990) or having big steps (Benferhat et al. 1999). There are finally close connections between possibility theory and

idempotent analysis (Maslov 1987; Kolokoltsov and Maslov 1997). The theory of large deviations in probability theory (Puhalskii 2001) also handles set functions that look like possibility measures (Nguyen and Bouchon-Meunier 2003). Here, we focus on the role of possibility theory in the theory of imprecise probability.

Possibility as Upper Probability

Let π be a possibility distribution where $\pi(s) \in [0,1]$. Let $\mathbf{P}(\pi)$ be the set of probability measures P such that $P \leq \Pi$, i.e., $\forall A \subseteq S, P(A) \leq \Pi(A)$. Then, the possibility measure Π coincides with the upper probability function P^* such that $P^*(A) = \sup\{P(A), P \in \mathbf{P}(\pi)\}$ while the necessity measure N is the lower probability function P_* such that $P_*(A) = \inf\{P(A), P \in \mathbf{P}(\pi)\}$; see Dubois and Prade (1992), De Cooman and Aeyels (1999) for details. P and π are said to be consistent if $P \in \mathbf{P}(\pi)$. The connection between possibility measures and imprecise probabilistic reasoning is especially promising for the efficient representation of nonparametric families of probability functions, and it makes sense even in the scope of modeling linguistic information (Walley and De Cooman 1999).

A possibility measure can be computed from a set of nested confidence subsets $\{A_1, A_2, \dots, A_m\}$ where $A_i \subset A_{i+1}$, $i = 1 \dots m-1$. Each confidence subset A_i is attached a positive confidence level λ_i interpreted as a lower bound of $P(A_i)$, hence a necessity degree. It is viewed as a certainty-qualified statement that generates a possibility distribution π_i as explained earlier in this paper. The corresponding possibility distribution is

$$\pi(s) = \min_{i=1, \dots, m} \pi_i(s) \\ = \begin{cases} 1 & \text{if } u \in A_1 \\ 1 - \lambda_{j-1} & \text{if } j = \max\{i : s \notin A_i\} > 1 \end{cases}$$

The information modeled by π can also be viewed as a nested random set $\{(A_i, v_i), i = 1, \dots, m\}$, where $v_i = \lambda_i - \lambda_{i-1}$. This framework allows for imprecision (reflected by the size of the A_i 's) and uncertainty (the v_i 's). And v_i is the probability that

the agent only knows that A_i contains the actual state (it is not $P(A_i)$). The random set view of possibility theory is well adapted to the idea of imprecise statistical data, as developed in (Gebhardt and Kruse 1993; Joslyn 1997). Namely, given a bunch of imprecise (not necessarily nested) observations (called focal sets), π supplies an approximate representation of the data, as $\pi(s) = \sum_{i:s \in A_i} v_i$.

The set $\mathbf{P}(\pi)$ contains many probability distributions, arguably too many. Neumaier (Neumaier 2004) has recently proposed a related framework, in a different terminology, for representing smaller subsets of probability measures using two possibility distributions instead of one. He basically uses a pair of distributions (δ, π) defined earlier in this paper he calls “cloud,” where δ is a guaranteed possibility distribution (in our terminology) such that $\pi \geq \delta$. A cloud models the (generally nonempty) set $\mathbf{P}(\pi) \cap \mathbf{P}(1 - \delta)$, viewing $1 - \delta$ as a standard possibility distribution.

Conditioning

There are two kinds of conditioning that can be envisaged upon the arrival of new information E . The first method presupposes that the new information alters the possibility distribution π by declaring all states outside E impossible. The conditional measure $\pi(\cdot | E)$ is such that $\Pi(B | E) \cdot \Pi(E) = \Pi(B \cap E)$. This is formally Dempster rule of conditioning of belief functions, specialized to possibility measures. The conditional possibility distribution representing the weighted set of confidence intervals is

$$\pi(s | E) = \begin{cases} \frac{\pi(s)}{\Pi(E)}, & \text{if } s \in E \\ 0 & \text{otherwise.} \end{cases}$$

De Baets et al. (1999) provide a mathematical justification of this notion in an infinite setting, as opposed to the min-based conditioning of qualitative possibility theory. Indeed, the maxitivity axiom extended to the infinite setting is not preserved by the min-based conditioning. The product-based conditioning leads to a notion of independence of the form $\Pi(B \cap E) =$

$\Pi(B) \cdot \Pi(E)$ whose properties are very similar to the ones of probabilistic independence (De Campos and Huete 1999).

Another form of conditioning (Dubois and Prade 1997; De Cooman 2001), more in line with the Bayesian tradition, considers that the possibility distribution π encodes imprecise statistical information, and event E only reflects a feature of the current situation, not of the state in general. Then, the value $\Pi(B | E) = \sup\{P(B | E), P(E) > 0, P \leq \Pi\}$ is the result of performing a sensitivity analysis of the usual conditional probability over $\mathbf{P}(\pi)$ (Walley 1991). Interestingly, the resulting set function is again a possibility measure, with distribution

$$\Pi(s | E) = \begin{cases} \max\left(\pi(s), \frac{\pi(s)}{\pi(s) + N(E)}\right), & \text{if } s \in E \\ 0 & \text{otherwise.} \end{cases}$$

It is generally less specific than π on E , as clear from the above expression, and becomes non-informative when $N(E) = 0$ (i.e., if there is no information about E). This is because $\pi(\cdot | E)$ is obtained from the focusing of the generic information π over the reference class E . On the contrary, $\pi(\cdot | E)$ operates a revision process on π due to additional knowledge asserting that states outside E are impossible. See De Cooman (2001) for a detailed study of this form of conditioning.

Probability-Possibility Transformations

The problem of transforming a possibility distribution into a probability distribution and conversely is meaningful in the scope of uncertainty combination with heterogeneous sources (some supplying statistical data, others linguistic data, for instance). It is useful to cast all pieces of information in the same framework. The basic requirement is to respect the consistency principle $\Pi \geq P$. The problem is then either to pick a probability measure in $\mathbf{P}(\pi)$ or to construct a possibility measure dominating P .

There are two basic approaches to possibility/probability transformations, which both respect a form of probability-possibility consistency. One,

due to Klir (Klir 1990; Geer and Klir 1992), is based on a principle of information invariance; the other (Dubois et al. 1993) is based on optimizing information content. Klir assumes that possibilistic and probabilistic information measures are commensurate. Namely, the choice between possibility and probability is then a mere matter of translation between languages “neither of which is weaker or stronger than the other” (quoting Klir and Parviz (1992)). It suggests that entropy and imprecision capture the same facet of uncertainty, albeit in different guises. The other approach, recalled here, considers that going from possibility to probability leads to increase the precision of the considered representation (as we go from a family of nested sets to a random element), while going the other way around means a loss of specificity.

From Possibility to Probability

The most basic example of transformation from possibility to probability is the Laplace principle of insufficient reason claiming that what is equally possible should be considered as equally probable. A generalized Laplacean indifference principle is then adopted in the general case of a possibility distribution π : the weights v_i bearing the sets A_i from the nested family of level cuts of π are uniformly distributed on the elements of these cuts A_i . Let P_i be the uniform probability measure on A_i . The resulting probability measure is $P = \sum_{i=1, \dots, m} v_i \cdot P_i$. This transformation, already proposed in 1982 (Dubois and Prade 1982), comes down to selecting the center of gravity of the set $\mathbf{P}(\pi)$ of probability distributions dominated by π . This transformation also coincides with Smets’ pignistic transformation (Smets 1990) and with the Shapley value of the “unanimity game” (another name of the necessity measure) in game theory. The rationale behind this transformation is to minimize arbitrariness by preserving the symmetry properties of the representation. This transformation from possibility to probability is one to one. Note that the definition of this transformation does not use the nestedness property of cuts of the possibility distribution. It applies all the same to non-nested random sets (or belief functions)

defined by pairs $\{(A_i, v_i), i = 1, \dots, m\}$, where v_i are non-negative reals such that $\sum_{i=1, \dots, m} v_i = 1$.

From Objective Probability to Possibility

From probability to possibility, the rationale of the transformation is not the same according to whether the probability distribution we start with is subjective or objective (Dubois et al. 2008). In the case of a statistically induced probability distribution, the rationale is to preserve as much information as possible. This is in line with the handling of Δ -qualified pieces of information representing observed evidence, considered earlier in this paper; hence, we select as the result of the transformation of a probability measure P the most specific possibility measure in the set of those dominating P (Dubois et al. 1993). This most specific element is generally unique if P induces a linear ordering on S . Suppose S is a finite set. The idea is to let $\Pi(A) = P(A)$, for these sets A having minimal probability among other sets having the same cardinality as A . If $p_1 > p_2 > \dots > p_n$, then $\Pi(A) = P(A)$ for sets A of the form $\{s_i, \dots, s_n\}$, and the possibility distribution is defined as $\pi_P(s_i) = \sum_{j=i, \dots, n} p_j$. Note that π_P is a kind of cumulative distribution of P . If there are equiprobable elements, the unicity of the transformation is preserved if equipossibility of the corresponding elements is enforced. Recently, this transformation was used to prove a rather surprising agreement between probabilistic indeterminateness as measured by Shannon entropy and possibilistic nonspecificity. Namely, it is possible to compare probability measures on finite sets in terms of their relative *peakedness* (a concept adapted from Birnbaum 1948) by comparing the relative specificity of their possibilistic transforms. Namely, let P and Q be two probability measures on S and π_P, π_Q the possibility distributions induced by our transformation. It can be proved that if $\pi_P \geq \pi_Q$ (i.e., P is less peaked than Q), then the Shannon entropy of P is higher than the one of Q (Dubois and Hüllermeier 2007). This result gives some grounds to the intuitions developed by Klir (1990), without assuming any commensurability between entropy and specificity indices.

Possibility Distributions Induced by Prediction Intervals

In the continuous case, moving from objective probability to possibility means adopting a representation of uncertainty in terms of prediction intervals around the mode viewed as the “most frequent value.” Extracting a prediction interval from a probability distribution or devising a probabilistic inequality can be viewed as moving from a probabilistic to a possibilistic representation. Namely, suppose a nonatomic probability measure P on the real line, with unimodal density p , and suppose one wishes to represent it by an interval I with a prescribed level of confidence $P(I) = \gamma$ of hitting it. The most natural choice is the most precise interval ensuring this level of confidence. It can be proved that this interval is of the form of a cut of the density, i.e., $I_\gamma = \{s, p(s) \geq \theta\}$ for some threshold θ . Moving the degree of confidence from 0 to 1 yields a nested family of prediction intervals that form a possibility distribution π consistent with P , the most specific one actually, having the same support and the same mode as P and defined by (Dubois et al. 1993):

$$\pi(\inf I_\gamma) = \pi(\sup I_\gamma) = 1 - \gamma = 1 - P(I_\gamma)$$

This kind of transformation again yields a kind of cumulative distribution according to the ordering induced by the density p . Similar constructs can be found in the statistical literature (Birnbaum 1948). More recently, Dubois et al. (Dubois et al. 2004b) noticed that starting from any family of nested sets around some characteristic point (the mean, the median, . . .), the above equation yields a possibility measure dominating P . Well-known inequalities of probability theory, such as those of Chebyshev and Camp-Meidel, can also be viewed as possibilistic approximations of probability functions. It turns out that for symmetric unimodal densities, each side of the optimal possibilistic transform is a convex function. Given such a probability density on a bounded interval $[a, b]$, the triangular fuzzy number whose core is the mode of p and the support is $[a, b]$ is thus a possibility distribution dominating

P regardless of its shape (and the tightest such distribution). These results justify the use of symmetric triangular fuzzy numbers as fuzzy counterparts to uniform probability distributions. They provide much tighter probability bounds than Chebyshev and Camp-Meidel inequalities for symmetric densities with bounded support. This setting is adapted to the modeling of sensor measurements (Mauris et al. 2000; Mauris 2013). These results are extended to more general distributions by Baudrit et al. (2004) and provide a tool for representing poor probabilistic information.

Subjective Possibility Distributions

The case of a subjective probability distribution is different. Indeed, the probability function is then supplied by an agent who is in some sense forced to express beliefs in this form due to rationality constraints and the setting of exchangeable bets. However, his actual knowledge may be far from justifying the use of a single well-defined probability distribution. For instance, in case of total ignorance about some value, apart from its belonging to an interval, the framework of exchangeable bets enforces a uniform probability distribution, on behalf of the principle of insufficient reason. Based on the setting of exchangeable bets, it is possible to define a subjectivist view of numerical possibility theory that differs from the proposal of Walley (1991). The approach developed by Dubois et al. (2003b) relies on the assumption that when an agent constructs a probability measure by assigning prices to lotteries, this probability measure is actually induced by a belief function representing the agent's actual state of knowledge. We assume that going from an underlying belief function to an elicited probability measure is achieved by means of the abovementioned pignistic transformation, changing focal sets into uniform probability distributions. The task is to reconstruct this underlying belief function under a minimal commitment assumption. In the paper (Dubois et al. 2003b), we pose and solve the problem of finding the least informative belief function having a given pignistic probability. We prove that it is unique and consonant, thus induced by a possibility distribution. This result exploits a simple partial

ordering between belief functions comparing their information content, in agreement with the expected cardinality of random sets. The obtained possibility distribution can be defined as the converse of the pignistic transformation (which is one to one for possibility distributions). It is subjective in the same sense as in the subjectivist school in probability theory. However, it is the least biased representation of the agent's state of knowledge compatible with the observed betting behavior. In particular, it is less specific than the one constructed from the prediction intervals of an objective probability. This transformation was first proposed in (Dubois and Prade 1983) for objective probability, interpreting the empirical necessity of an event as summing the excess of probabilities of realizations of this event with respect to the probability of the most likely realization of the opposite event.

Mean Values in Possibility Theory

Possibilistic mean values can be defined using Choquet integrals with respect to possibility and necessity measures (Dubois and Prade 1985; De Cooman 2001) and come close to defuzzification methods (Van Leekwijck and Kerre 2001). A fuzzy interval is a fuzzy set of reals whose membership function is unimodal and upper semi-continuous. Its α -cuts are closed intervals. Interpreting a fuzzy interval M , associated to a possibility distribution μ_M , as a family of probabilities, upper and lower mean values $E^*(M)$ and $E_*(M)$, can be defined as (Dubois and Prade 1987)

$$E_*(M) = \int_0^1 \inf M_\alpha d\alpha; \quad E^*(M) = \int_0^1 \sup M_\alpha d\alpha$$

where M_α is the α -cut of M .

Then, the mean interval $E(M) = [E_*(M), E^*(M)]$ of M is the interval containing the mean values of all random variables consistent with M , that is, $E(M) = \{E(P) \mid P \in \mathbf{P}(\mu_M)\}$, where $E(P)$ represents the expected value associated to the probability measure P . That the "mean value" of a fuzzy interval is an interval seems to be intuitively satisfactory. Particularly the mean interval of a (regular) interval $[a, b]$ is this interval itself.

The upper and lower mean values are linear with respect to the addition of fuzzy numbers. Define the addition $M + N$ as the fuzzy interval whose cuts are $M_\alpha + N_\alpha = \{s + t, s \in M_\alpha, t \in N_\alpha\}$ defined according to the rules of interval analysis. Then, $E(M + N) = E(M) + E(N)$, and similarly for the scalar multiplication $E(aM) = aE(M)$, where aM has membership grades of the form $\mu_M(s/a)$ for $a \neq 0$. In view of this property, it seems that the most natural defuzzification method is the middle point $\hat{E}(M)$ of the mean interval (originally proposed by Yager 1981). Other defuzzification techniques do not generally possess this kind of linearity property. $\hat{E}(M)$ has a natural interpretation in terms of simulation of a fuzzy variable (Chanas and Nowakowski 1988) and is the mean value of the pignistic transformation of M . Indeed, it is the mean value of the empirical probability distribution obtained by the random process defined by picking an element α in the unit interval at random and then an element s in the cut M_α at random.

Applications and Future Directions

Possibility theory has not been the main framework for engineering applications of fuzzy sets in the past. However, on the basis of its connections to symbolic artificial intelligence, to decision theory, and to imprecise statistics, we consider that it has significant potential for further applied developments in a number of areas, including some where fuzzy sets are not yet always accepted. Only some directions are pointed out here:

1. Rules with exceptions can be modeled by means of conditional possibility (Benferhat et al. 1997), based on its capability to account for nonmonotonic inference, as shown in section “Qualitative Possibility Theory”. Possibility theory has also enabled a typology of fuzzy rules to be laid bare, distinguishing rules whose purpose is to propagate uncertainty through reasoning steps, from rules whose main purpose is similarity-based interpolation (Dubois and Prade 1996), depending on the choice of a many-valued implication connective that models a rule. The bipolar view of information based on (δ, π) pairs sheds new light on the debate between conjunctive and implicative representation of rules (Dubois et al. 2003c). Representing a rule as a material implication focuses on counterexamples to rules, while using a conjunction between antecedent and consequent points out examples of the rule and highlights its positive content. Traditionally in fuzzy control and modeling, the latter representation is adopted, while the former is the logical tradition. Introducing fuzzy implicative rules in modeling accounts for constraints or landmark points the model should comply with (as opposed to observed data) (Galichet et al. 2004). The bipolar view of rules in terms of examples and counterexamples may turn out to be very useful when extracting fuzzy rules from data (Dubois et al. 2003d).
2. Possibility theory also provides an approach to preference modeling; see, e.g., (Dubois et al. 2013). In constraint-directed reasoning, both prioritized and soft constraints can be captured by possibility distributions expressing degrees of feasibility rather than plausibility (Dubois et al. 1996). Possibility offers a natural setting for fuzzy optimization whose aim is to balance the levels of satisfaction of multiple fuzzy constraints (instead of minimizing an overall cost) (Dubois and Fortemps 1999). Qualitative decision criteria are particularly adapted to the handling of uncertainty in this setting. Applications of possibility theory-based decision-making can be found in scheduling (Dubois et al. 1995; Slowinski and Hapke 2000; Chanas and Zielinski 2001; Chanas et al. 2002). In the same spirit, let us also point out applications to database querying (Bosc and Prade 1997; Bosc et al. 2010) and to the handling of databases with uncertain data (Bosc et al. 2009).
3. Quantitative possibility theory is the natural setting for a reconciliation between probability and fuzzy sets. An important research direction is the comparison between fuzzy interval analysis (Dubois et al. 2000d) and random variable calculations with a view to unifying them

(Dubois and Prade 1991b). Indeed, a current major concern, in, for instance, risk analysis studies, is to perform uncertainty propagation under poor data and without independence assumptions (see the papers in the special issue Helton and Oberkampf 2004). Finding the potential of possibilistic representations in computing conservative bounds for such probabilistic calculations is certainly a major challenge (Guyonnet et al. 2003). Methods for joint propagation of possibilistic and probabilistic information have been devised (Baudrit et al. 2006), based on casting both in a random set setting (Baudrit et al. 2007). Indeed, the active area of fuzzy random variables is also connected to this question (Gil 2001). The introduction of possibility theory in regression analysis with fuzzy data, first proposed in (Tanaka and Guo 1999), comes down to an epistemic view of fuzzy data whereby one tries to construct the envelope of all linear regression results that could have been obtained, had the data been precise. This view has been applied to the kriging problem in geostatistics (Loquin and Dubois 2012). Another use of possibility theory consists in exploiting possibility-probability transforms to develop a form of quantile regression on crisp data (Serrurier and Prade 2007, 2013), yielding a fuzzy function that is much more faithful to the data set than what a fuzzified linear function can offer. Other applications of possibility theory to data analysis can be found in (Wolkenhauer 1998; Tanaka and Guo 1999; Borgelt et al. 2000b).

4. One might also mention the well-known possibilistic clustering technique (Krishnapuram and Keller 1993; Bezdek et al. 1999). However, it is only loosely related to possibility theory. This name is due to the use of fuzzy clusters with (almost) unrestricted membership functions that no longer form a usual fuzzy partition. But one might use it to generate genuine possibility distributions, where possibility derives from similarity, a point of view already mentioned before.

Other applications of possibility theory can be found in fields such as diagnosis (Cayrac et al. 1996; Boverie et al. 2002), belief revision (Benferhat et al. 2002b), argumentation (Amgoud and Prade 2004, 2009), or case-based reasoning (Dubois et al. 2002; Hüllermeier et al. 2002).

Lastly, possibility theory is also being studied from the point of view of its relevance in cognitive psychology. Experimental results (Raufaste et al. 2003) suggest that there are situations where people reason about uncertainty using the rules of possibility theory rather than with those of probability theory.

Bibliography

- Amgoud L, Prade H (2004) Reaching agreement through argumentation: a possibilistic approach. In: Proceedings of the 9th international conference on principles of knowledge representation and reasoning (KR'04), Whistler. AAAI Press, pp 175–182
- Amgoud L, Prade H (2009) Using arguments for making and explaining decisions. *Artif Intell* 173:413–436
- Banerjee M, Dubois D (2009) A simple modal logic for reasoning about revealed beliefs. In: Sossai C, Chemello G (eds) Proceedings of the 10th European conference symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU09), Verona, 1–3 July 2009. LNCS, vol 5590, 805816. Springer, Berlin
- Banerjee M, Dubois D (2014) A simple logic for reasoning about incomplete knowledge. *Int J of Approx Reason* 55:639–653
- Baudrit C, Dubois D, Fargier H (2004) Practical representation of incomplete probabilistic information. In: Lopez-Diaz M et al (eds) Soft methods in probability and statistics (Proceedings of the 2nd international conference, Oviedo, Spain). Springer, Berlin, pp 149–156
- Baudrit C, Guyonnet D, Dubois D (2006) Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans Fuzzy Syst* 14:593–608
- Baudrit C, Couso I, Dubois D (2007) Joint propagation of probability and possibility in risk analysis: towards a formal framework. *Int J of Approx Reason* 45:82–105
- Ben Amor N, Benferhat S (2005) Graphoid properties of qualitative possibilistic independence relations. *Int J Unc Fuzz Knowl Based Syst* 13:59–97
- Ben Amor N, Mellouli K, Benferhat S, Dubois D, Prade H (2002) A theoretical framework for possibilistic independence in a weakly ordered setting. *Int J Unc Fuzz Knowl Based Syst* 10:117–155

- Ben Amor N, Benferhat S, Mellouli K (2003) Anytime propagation algorithm for min-based possibilistic graphs. *Soft Comput* 8(2):150–161
- Benferhat S, Dubois D, Prade H (1997) Nonmonotonic reasoning, conditional objects and possibility theory. *Artif Intell* 92:259–276
- Benferhat S, Dubois D, Prade H (1998) Practical handling of exception-tainted rules and independence information in possibilistic logic. *Appl Intell* 9:101–127
- Benferhat S, Dubois D, Prade H (1999) Possibilistic and standard probabilistic semantics of conditional knowledge bases. *J Logic Comput* 9:873–895
- Benferhat S, Dubois D, Garcia L, Prade H (2002a) On the transformation between possibilistic logic bases and possibilistic causal networks. *Int J Approx Reason* 29:135–173
- Benferhat S, Dubois D, Prade H, Williams M-A (2002b) A practical approach to revising prioritized knowledge bases. *Stud Logic* 70:105–130
- Benferhat S, Dubois D, Garcia L, Prade H (2002c) On the transformation between possibilistic logic bases and possibilistic causal networks. *Int J Approx Reason* 29(2):135–173
- Bezdek J, Keller J, Krishnapuram R, Pal N (1999) Fuzzy models and algorithms for pattern recognition and image processing, The handbooks of fuzzy sets series. Kluwer, Boston
- Birnbaum ZW (1948) On random variables with comparable peakedness. *Ann Math Stat* 19:76–81
- Borgelt C, Kruse R (2003) Operations and evaluation measures for learning possibilistic graphical models. *Artif Intell* 148(1–2):385–418
- Borgelt C, Gebhardt J, Kruse R (2000a) Possibilistic graphical models. In: Della Riccia G et al (eds) Computational intelligence in data mining. Springer, Wien, pp 51–68
- Borgelt C, Gebhardt J, Kruse R (2000b) Possibilistic graphical models. In: Della Riccia G et al (eds) Computational intelligence in data mining, CISM series, N408. Springer, Berlin
- Bosc P, Prade H (1997) An introduction to the fuzzy set and possibility theory-based treatment of soft queries and uncertain of imprecise databases. In: Smets P, Motro A (eds) Uncertainty management in information systems. Kluwer, Dordrecht, pp 285–324
- Bosc P, Pivert O, Prade H (2009) A model based on possibilistic certainty levels for incomplete databases. In: Godo L, Pugliese A (eds) Proceedings of the 3rd international conference on scalable uncertainty management (SUM 2009), Washington, DC, 28–30 Sept 2009. LNCS, vol 5785. Springer, pp 80–94
- Bosc P, Pivert O, Prade H (2010) A possibilistic logic view of preference queries to an uncertain database. In: Proceedings of the 19th IEEE international conference on fuzzy systems (FUZZ-IEEE10), 581595
- Boverie S, Dubois D, Gurandel X, De Mouzon O, Prade H (2002) Online diagnosis of engine dyno test benches: a possibilistic approach. In: Proceedings of the 15th European conference on artificial intelligence. IOS Press, Amsterdam, pp 658–662
- Cayrac D, Dubois D, Prade H (1996) Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. *IEEE Trans Fuzzy Syst* 4:251–269
- Chanas S, Nowakowski M (1988) Single value simulation of fuzzy variable. *Fuzzy Set Syst* 25:43–57
- Chanas S, Zielinski P (2001) Critical path analysis in the network with fuzzy activity times. *Fuzzy Set Syst* 122:195–204
- Chanas S, Dubois D, Zielinski P (2002) Necessary criticality in the network with imprecise activity times. *IEEE Trans Man Machine Cybernet* 32:393–407
- Cohen LJ (1977) The probable and the provable. Clarendon, Oxford
- De Baets B, Tsiporkova E, Mesiar R (1999) Conditioning in possibility with strict order norms. *Fuzzy Set Syst* 106:221–229
- De Campos LM, Huete JF (1999) Independence concepts in possibility theory. *Fuzzy Sets Syst* 103:127–152, 487–506
- De Cooman G (1997) Possibility theory. Part I: measure- and integral-theoretic groundwork; Part II: conditional possibility; Part III: possibilistic independence. *Int J General Syst* 25:291–371
- De Cooman G (2001) Integration and conditioning in numerical possibility theory. *Annal Math Artif Intell* 32:87–123
- De Cooman G, Aeyels D (1999) Supremum-preserving upper probabilities. *Inform Sci* 118:173–212
- Dubois D (1986) Belief structures, possibility theory and decomposable measures on finite sets. *Comput Artif Intell* 5:403–416
- Dubois D (2006) Possibility theory and statistical reasoning. *Comput Stat Data Anal* 51(1):47–69
- Dubois D, Fargier H (2003) Qualitative decision rules under uncertainty. In: Della Riccia G et al (eds) Planning based on decision theory, CISM Courses and Lectures 472. Springer, Wien, pp 3–26
- Dubois D, Fortemps P (1999) Computing improved optimal solutions to maxmin flexible constraint satisfaction problems. *Eur J Oper Res* 118:95–126
- Dubois D, Hüllermeier E (2007) Comparing probability measures using possibility theory: a notion of relative peakedness. *Inter J Approx Reason* 45:364–385
- Dubois D, Prade H (1980) Fuzzy sets and systems: theory and applications. Academic, New York
- Dubois D, Prade H (1982) On several representations of an uncertain body of evidence. In: Gupta M, Sanchez E (eds) Fuzzy information and decision processes. North-Holland, Amsterdam, pp 167–181
- Dubois D, Prade H (1983) Unfair coins and necessity measures: a possibilistic interpretation of histograms. *Fuzzy Set Syst* 10(1):15–20
- Dubois D, Prade H (1985) Evidence measures based on fuzzy information. *Automatica* 21:547–562
- Dubois D, Prade H (1987) The mean value of a fuzzy number. *Fuzzy Set Syst* 24:279–300

- Dubois D, Prade H (1988) Possibility theory. Plenum, New York
- Dubois D, Prade H (1991a) Epistemic entrenchment and possibilistic logic. *Artif Intell* 50:223–239
- Dubois D, Prade H (1991b) Random sets and fuzzy interval analysis. *Fuzzy Set Syst* 42:87–101
- Dubois D, Prade H (1992) When upper probabilities are possibility measures. *Fuzzy Set Syst* 49:65–74
- Dubois D, Prade H (1996) What are fuzzy rules and how to use them. *Fuzzy Set Syst* 84:169–185
- Dubois D, Prade H (1997) Bayesian conditioning in possibility theory. *Fuzzy Set Syst* 92:223–240
- Dubois D, Prade H (1998) Possibility theory: qualitative and quantitative aspects. In: Gabbay DM, Smets P (eds) *Handbook of defeasible reasoning and uncertainty management systems*, vol 1. Kluwer, Dordrecht, pp 169–226
- Dubois D, Prade H (2004) Possibilistic logic: a retrospective and prospective view. *Fuzzy Set Syst* 144:3–23
- Dubois D, Prade H (2007) Toward multiple-agent extensions of possibilistic logic. In: *Proceedings of the IEEE international conference on fuzzy systems (FUZZ-IEEE 2007)*, London, UK, pp 187–192
- Dubois D, Prade H (2015) Possibilistic logic – An overview. In: *Handbook of the History of Logic. Computational Logic*. (J. Siekmann, vol. ed.; D. M. Gabbay, J. Woods, series eds.), Elsevier, Amsterdam 9:283–342
- Dubois D, Prade H, Sandri S (1993) On possibility/probability transformations. In: Lowen R, Roubens M (eds) *Fuzzy logic: state of the art*. Kluwer, Dordrecht, pp 103–112
- Dubois D, Lang J, Prade H (1994) Possibilistic logic. In: Gabbay DM et al (eds) *Handbook of logic in AI and logic programming*, vol 3. Oxford University Press, New York, pp 439–513
- Dubois D, Fargier H, Prade H (1995) Fuzzy constraints in job-shop scheduling. *J Intell Manufact* 6:215–234
- Dubois D, Fargier H, Prade H (1996) Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty. *Appl Intell* 6:287–309
- Dubois D, Fariñas del Cerro L, Herzig A, Prade H (1997) Qualitative relevance and independence: a roadmap. In: *Proceedings of the 15th international joint conference on artificial intelligence*, Nagoya, pp 62–67
- Dubois D, Moral S, Prade H (1997b) A semantics for possibility theory based on likelihoods. *J Math Anal Appl* 205:359–380
- Dubois D, Nguyen HT, Prade H (2000a) Fuzzy sets and probability: misunderstandings, bridges and gaps. In: Dubois D, Prade H (eds) *Fundamentals of fuzzy sets*. Kluwer, Boston, pp 343–438
- Dubois D, Hajek P, Prade H (2000b) Knowledge-driven versus data-driven logics. *J Logic Lang Inform* 9:65–89
- Dubois D, Prade H, Sabbadin R (2000c) Qualitative decision theory with Sugeno integrals. In: Grabisch M, Murofushi T, Sugeno M (eds) *Fuzzy measures and integrals theory and applications*. Physica-Verlag, Heidelberg, pp 314–322
- Dubois D, Kerre E, Mesiar R, Prade H (2000d) Fuzzy interval analysis. In: Dubois D, Prade H (eds) *Fundamentals of fuzzy sets*. Kluwer, Boston, pp 483–581
- Dubois D, Prade H, Sabbadin R (2001) Decision-theoretic foundations of possibility theory. *Eur J Oper Res* 128: 459–478
- Dubois D, Hüllermeier E, Prade H (2002) Fuzzy set-based methods in instance-based reasoning. *IEEE Trans Fuzzy Syst* 10:322–332
- Dubois D, Fargier H, Perny P (2003a) Qualitative decision theory with preference relations and comparative uncertainty: an axiomatic approach. *Artif Intell* 148: 219–260
- Dubois D, Prade H, Smets P (2003b) A definition of subjective possibility. *Badania Operacyjne i Decyzje* (Wrocław) 4:7–22
- Dubois D, Prade H, Ughetto L (2003c) A new perspective on reasoning with fuzzy rules. *Int J Intell Syst* 18: 541–567
- Dubois D, Hüllermeier E, Prade H (2003d) A note on quality measures for fuzzy association rules. In: De Baets B, Bilgic T (eds) *Fuzzy sets and systems* (Proceedings of the 10th international fuzzy systems association World Congress IFSA 2003, Istanbul, Turkey), LNAI 2715. Springer, Berlin, pp 346–353
- Dubois D, Fargier H, Prade H (2004a) Ordinal and probabilistic representations of acceptance. *J Artif Intell Res* 22:23–56
- Dubois D, Foulloy L, Mauris G, Prade H (2004b) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Comput* 10:273–297
- Dubois D, Prade H, Smets P (2008) A definition of subjective possibility. *Inte J Approx Reason* 48:352–364
- Dubois D, Prade H, Schockaert S (2012) Stable models in generalized possibilistic logic. In: Brewka G, Eiter T, McIlraith SA (eds) *Proceedings of the 13th international conference on principles of knowledge representation and reasoning (KR12)*, Roma, June 10–14, AAAI Press, pp 519–529
- Dubois D, Prade H, Rico A (2013) Qualitative capacities as imprecise possibilities. In: van der Gaag LC (ed) *Proceedings of the 12th European conference on symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU'13)*. LNAI, vol 7958. Springer, pp 169–180
- Dubois D, Prade H, Touazi F (2013) Conditional preference nets and possibilistic logic. In: van der Gaag LC (ed) *Proceedings of the 12th European conference on symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU'13)*, LNAI 7958. Springer, Berlin, pp 181–193
- Fargier H, Sabbadin R (2005) Qualitative decision under uncertainty: Back to expected utility. *textit. Artif Intell* 164:245–280
- Gaines BR, Kohout L (1975) Possible automata. In: *Proceedings of the international symposium multiple-valued logics*, Bloomington, pp 183–196

- Galichet S, Dubois D, Prade H (2004) Imprecise specification of ill-known functions using gradual rules. *Int J Approx Reason* 35:205–222
- Gärdenfors P (1998) *Knowledge in flux*. MIT Press, Cambridge, MA
- Gebhardt J, Kruse R (1993) The context model I. *Int J Approx Reason* 9:283–314
- Geer JF, Klir GJ (1992) A mathematical analysis of information-preserving transformations between probabilistic and possibilistic formulations of uncertainty. *Int J of General Syst* 20:143–176
- Gil M (ed) (2001) Fuzzy random variables, special issue. *Inform Sci*, 133:1–100
- Grabisch M, Murofushi T, Sugeno M (eds) (2000) *Fuzzy measures and integrals theory and applications*. Physica-Verlag, Heidelberg
- Guyonnet D, Bourgine B, Dubois D, Fargier H, Come B, Chiles J-P (2003) Hybrid approach for addressing uncertainty in risk assessments. *J Environ Eng* 129: 68–78
- Helton JC, Oberkampf WL (eds) (2004) Alternative representations of uncertainty. In: *Reliability engineering and systems safety*, vol 85. Elsevier, 369 p
- Hughes GE, Cresswell MJ (1968) *An introduction to modal logic*. Methuen, London
- Hüllermeier E, Dubois D, Prade H (2002) Model adaptation in possibilistic instance-based reasoning. *IEEE Trans Fuzzy Syst* 10:333–339
- Joslyn C (1997) Measurement of possibilistic histograms from interval data. *Int J of General Syst* 26:9–33
- Klir GJ (1990) A principle of uncertainty and information invariance. *Int J of General Syst* 17:249–275
- Klir GJ, Folger T (1988) *Fuzzy sets, uncertainty and information*. Prentice Hall, Englewood Cliffs
- Klir GJ, Parviz B (1992) Probability-possibility transformations: a comparison. *Int J General Syst* 21:291–310
- Kolokoltsov VN, Maslov VP (1997) *Idempotent analysis and applications*. Kluwer, Dordrecht
- Krishnapuram R, Keller J (1993) A possibilistic approach to clustering. *IEEE Trans Fuzzy Syst* 1:98–110
- Lehmann D, Magidor M (1992) What does a conditional knowledge base entail? *Artif Intell* 55:1–60
- Lewis DL (1973) *Counterfactuals*. Basil Blackwell, Oxford
- Loquin K, Dubois D (2012) A fuzzy interval analysis approach to kriging with ill-known variogram and data. *Soft Comput* 16:769–784
- Maslov V (1987) *Méthodes opératoires*. Mir Publications, Moscow
- Mauris G (2013) A Review of Relationships Between Possibility and Probability Representations of Uncertainty in Measurement. *IEEE Trans Measure Instrum* 62(3):622–632
- Mauris G, Lasserre V, Foulloy L (2000) Fuzzy modeling of measurement data acquired from physical sensors. *IEEE Trans Measure Instrum* 49:1201–1205
- Neumaier A (2004) Clouds, fuzzy sets and probability intervals. *Reliable Comput* 10:249–272
- Nguyen HT, Bouchon-Meunier B (2003) Random sets and large deviations principle as a foundation for possibility measures. *Soft Comput* 8:61–70
- Pearl J (1990) System Z: a natural ordering of defaults with tractable applications to default reasoning. In: *Proceeding of the 3rd conference theoretical aspects of reasoning about knowledge*. Morgan Kaufmann, San Francisco, pp 121–135
- Puhalskii A (2001) *Large deviations and idempotent probability*. Chapman and Hall, New York
- Raufaste E, Da Silva Neves R, Marine C (2003) Marine testing the descriptive validity of possibility theory in human judgements of uncertainty. *Artif Intell* 148: 197–218
- Savage LJ (1972) *The foundations of statistics*. Dover, New York
- Serrurier M, Prade H (2007) A general framework for imprecise regression. In: *Proceedings of the IEEE international conference on fuzzy systems (FUZZ-IEEE'07)*, London, July 23–26, pp 1597–1602
- Serrurier M, Prade H (2013) An informational distance for estimating the faithfulness of a possibility distribution, viewed as a family of probability distributions, with respect to data. *Int J Approx Reason* 54:919–933
- Shackle GLS (1961) *Decision, order and time in human affairs*, 2nd edn. Cambridge University Press, UK
- Shafer G (1987) Belief functions and possibility measures. In: Bezdek JC (ed) *Analysis of fuzzy information vol I: mathematics and logic*. CRC Press, Boca Raton, pp 51–84
- Slowinski R, Hapke M (eds) (2000) *Scheduling under fuzziness*. Physica-Verlag, Heidelberg
- Smets P (1990) Constructing the pignistic probability function in a context of uncertainty. In: Henrion M et al (eds) *Uncertainty in artificial intelligence*, vol 5. North-Holland, Amsterdam, pp 29–39
- Spohn W (1990) A general, nonprobabilistic theory of inductive reasoning. In: Shachter RD et al (eds) *Uncertainty in artificial intelligence*, vol 4. North Holland, Amsterdam, pp 149–158
- Sudkamp T (2002) Similarity and the measurement of possibility. *Actes Rencontres Francophones sur la Logique Floue et ses Applications* (Montpellier, France). Cepadues Editions, Toulouse, pp 13–26
- Tanaka H, Guo PJ (1999) Possibilistic data analysis for operations research. Physica-Verlag, Heidelberg
- Van Leekwijck W, Kerre EE (2001) Defuzzification: criteria and classification. *Fuzzy Set Syst* 108:303–314
- Walley P (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall, New York
- Walley P (1996) Measures of uncertainty in expert systems. *Artif Intell* 83(1-58):1996
- Walley P, De Cooman G (1999) A behavioural model for linguistic uncertainty. *Information Sciences* 134:1–37
- Whalen T (1984) Decision making under uncertainty with various assumptions about available information. *IEEE Trans Syst Man Cybernet* 14:888–900
- Wolkenhauer O (1998) *Possibility theory with applications to data analysis*. Research Studies Press, Chichester

- Yager RR (1979) Possibilistic decision making. *IEEE Trans Syst Man Cybernet* 9:388–392
- Yager RR (1981) A procedure for ordering fuzzy subsets of the unit interval. *Inform Sci* 24:143–161
- Yager RR (1983) An introduction to applications of possibility theory. *Hum Syst Manage* 3:246–269
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Set Syst* 1:3–28
- Zadeh LA (1979) Fuzzy sets and information granularity. In: Gupta MM, Ragade R, Yager RR (eds) *Advances in fuzzy set theory and applications*. North-Holland, Amsterdam, pp 3–18
- Zadeh LA (1982) Possibility theory and soft data analysis. In: Cobb L, Thrall R (eds) *Mathematical frontiers of social and policy sciences*. Westview Press, Boulder, pp 69–129

Rough Sets: Foundations and Perspectives

James F. Peters¹, Andrzej Skowron² and Jaroslaw Stepaniuk³

¹Computational Intelligence Laboratory, University of Manitoba, Winnipeg, MB, Canada

²Institute of Mathematics, Warsaw University, Warsaw, Poland

³Computer Science, Bialystok University of Technology, Bialystok, Poland

Article Outline

Glossary

Definition of the Subject

Introduction

Approximation Spaces

Rough Sets

Dimensionality Reduction

Summary

Future Directions

Bibliography

Glossary

Approximation The replacement of mathematical objects by others that resemble them in certain respects (Vinogradov 1995).

Approximation space An approximation space is denoted by $(O, \mathcal{F}, \sim_B)$, where O is a set of perceived objects, $\mathcal{F}$ is a set of probe functions representing object features, and $\sim_B$ is an indiscernibility relation defined relative to $B \subseteq \mathcal{F}$. This approximation space is considered fundamental because it provided a framework for the original rough set theory (Pawlak 1981; Pawlak and Skowron 2007). Several generalizations of this definition of approximation space have been proposed (see, e. g., Pawlak and Skowron 2007; Peters et al. 2007; Skowron and Stepaniuk 1996; Skowron et al.

2005, 2006; Stepaniuk 2000, 2007; Ziarko 1993).

Attribute A quality regarded as characteristic or inherent in an object (Murray et al. 1933). In rough set theory, an attribute a of an object x is represented by a partial function $f_a(x) = v$, where v is a value in the range of f_a . In rough set theory, the function f_a is often called an attribute (Pawlak 1991; Polkowski 2002).

Boundary region The B -boundary region of an approximation of a set X is denoted by $\text{Bnd}_B X$ and is defined relative to a set of functions B representing features of objects in X as well as the lower approximation B_*X and the upper approximation B^*X , where

$$\text{Bnd}_B X = B^*X \setminus B_*X = \{x \mid x \in B^*X \text{ and } x \notin B_*X\}.$$

Elementary set A B -class in the quotient set $X/\sim_B$.

Equivalence class Given a relation $\sim$, an equivalence class is a set denoted by $[x]$ or $[x]_\sim$ (Devlin 1993) in the quotient set $X/\sim$ (See Glossary item “Quotient set”), where

$$[x] = \{x' \in X \mid x \sim x'\}.$$

Equivalence relation A reflexive, symmetric and transitive relation $\sim \subseteq X \times X$. An equivalence relation $\sim$ on a set X defines a partition of X into classes.

Feature Make, form, fashion, shape (of an object) (Murray et al. 1933). A characteristic of an object perceived by the senses or knowable by the mind (Peters 2008; Skowron and Peters 2008). In rough set theory, a feature f of an object x is represented by a function $\phi_f(x) = v$, where v is a value in the range of ϕ_f (e. g., $\phi_{\sharp}(x)$ as a measure of the tonality $\sharp$ feature of a Chopin Mazurka x) (Peters 2008; Skowron and Peters 2008). The function ϕ_f is sometimes also termed an attribute (Pawlak 1991; Polkowski 2002).

Indiscernibility relation An equivalence relation

$$\sim_B = \{(x, x') \in X \times X \mid f(x) = f(x') \text{ for any } f \in B\},$$

where X denotes a set of objects, B denotes a set of functions, and $f \in B$ is a function representing a feature of an object $x \in X$. The notation used to denote an equivalence relation in rough set theory has varied widely over time. For example, $\tilde{B}$ was originally introduced by Zdzisław Pawlak in 1981 (Pawlak 1981). Later, $\text{Ind}(B)$ (Grzymala-Busse and Grzymala-Busse 2007; Nguyen 2006; Pawlak 1991; Wojna 2007) or IND_B (Polkowski 2002) or Ind_B (Gomolinska 2005) or IND (Wojna 2007) or $I(B)$ (Pawlak and Skowron 2007) or $=_B$ (Greco et al. 2005) has also been used to denote an equivalence relation on a set X defined relative to attributes of objects. In rough set theory, the equivalence relation $\sim_B$ was introduced by Zdzisław Pawlak (1981).

Information granule Information granules are obtained in the process of granulation. Granulation can be viewed as a human way of achieving data compression and it plays a key role in implementing the divide-and-conquer strategy in human problem-solving. An *information granule* represents a set of objects that have descriptions matching the granule (Skowron and Peters 2008), e. g. elementary set $[x]_B$, lower approximation B_*X , quotient set $X/\sim_B$.

Lower approximation The B -lower approximation of a set X is denoted by B_*X and is defined relative to a set of functions B representing features of objects in X and the quotient set $X/\sim_B$, where

$$B_*X = \bigcup_{x:[x]_B \subseteq X} [x]_B.$$

Object Something perceptible to the senses or knowable by the mind (Murray et al. 1933).

Information Whatever is conveyed or represented by a particular sequence of symbols (Murray et al. 1933). In rough set theory, information is derivable either from the patterns in a particular information table or from what can be observed in a particular approximation space (Pawlak 1981; Pawlak and Skowron 2007).

Information system A system to represent knowledge (Konrad et al. 1981; Pawlak 1973;

Pawlak and Skowron 2007). Syntactic representation of knowledge in table form (Konrad et al. 1981; Polkowski 2002).

Partition of a non-empty set X A family of non-empty, pairwise disjoint subsets of X (called classes) such that the union of this family is equal to X .

Quotient set Set of all classes in a partition defined by an equivalence relation $\sim$ on a set X (denoted by $X/\sim$).

Rough set A set X is considered a rough set if, and only if it has a non-empty B -boundary $\text{Bnd}_B X$, i. e., the B -approximation of X has a non-empty boundary.

Upper approximation The B -upper approximation of X is denoted by B^*X and is defined relative to a set of functions B representing features of objects in X and the quotient set $X/\sim_B$, where

$$B^*X = \bigcup_{x:[x]_B \cap X \neq \emptyset} [x]_B.$$

Definition of the Subject

Rough set theory was introduced by Zdzisław Pawlak (1926–2006) during the early 1980s. This theory has two distinguishing hallmarks: knowledge description systems introduced by Pawlak in 1973 (Pawlak 1973) and set approximation (See Glossary Items “Approximation”, “Boundary region”, “Lower approximation”, “Upper approximation”) as a means of classifying a set (Pawlak 1981). Knowledge description systems represent our knowledge about sample objects in tabular form. This knowledge results from identifying a set of attributes A (apparent qualities) of the objects. A total function $\rho : X \times A \rightarrow V_a$ maps an attribute $a \in A$ of an object $x \in X$ to a value $v \in V_a$. Our knowledge about objects $x_1, \dots, x_n$ having attributes $a_1, \dots, a_k$ then can be represented in table form as rows of knowledge descriptions, e. g., description of x_i as a tuple.

$$(\rho(x_i, a_1), \rho(x_i, a_2), \dots, \rho(x_i, a_k)).$$

At this point, there is a natural transition to the formulation of an indiscernibility relation $\sim_B$ that

defines a partition of a set of sample objects relative to set $B \subseteq A$ (see Glossary item “Indiscernibility relation”). In many ways, this relation is both fecund and important. The fecundity (high fertility) of this relation can be seen in several ways. First, $\sim_B$ makes it possible to organize tabular representations of objects into equivalence classes. This, of course, means that now our observations can be made at the level of classes (elementary granules) instead of the more intransigent level of individual objects. Second, the majesty of $\sim_B$ can be seen in Pawlak’s discovery of lower and upper descriptions of objects. These descriptions accrue naturally from lower and upper approximations of a set of objects. Third, it is then possible gauge the *roughness* of our knowledge about a set of objects by considering the size of the boundary of an approximation. The roughness of our knowledge about a set of sample objects is directly proportional to the size of the approximation boundary. So, in effect, the indiscernibility relation lead to the discovery of rough sets.

The importance of the indiscernibility relation can be seen in a number of ways, if one considers the ordinary difficulties associated with the complexity of ordinary, perceptual objects. The introduction of this relation led to the creation of an approximation space that provides a framework for perception or observation on the level of classes (Orłowska 1985a; Pawlak 1981). An important byproduct of the approach to approximation in rough set theory is information granulation (Skowron and Peters 2008). Each selection of features of a set of sample objects leads to granulation of the information associated with the objects. For a set of sample objects X with selected features represented by a set of functions B , a rich harvest of information granules from the partition of X defined by the indiscernibility relation $\sim_B$, e. g., individual classes $[x]_B$, quotient space $X/\sim_B$, lower approximation B_*X , upper approximation B^*X , and boundary $\text{Bnd}_B X$. From the introduction of knowledge description systems and approximation of our knowledge springs various forms of learning, starting with learning from examples exemplified by a number rough set toolsets

released during the past decade (see, e. g., Int n.d.; Rough Set Database n.d.).

Basic ideas of rough set theory and its extensions, as well as many interesting applications can be found in numerous books (Rough Set Database n.d.), issues of the Transactions on Rough Sets (n.d.), special issues of other journals, proceedings of international conferences, tutorials or surveys (see, references and further readings listed in this paper and, e. g., (Pawlak and Skowron 2007) and numerous web pages such as (Int n.d.; University of Regina Electronic Bulletin of the Rough Set Community n.d; Warsaw University Logic Group n.d).

Introduction

This chapter gives a concise overview of some of the features of the foundations and perspectives of rough set theory. The foundations of rough set theory are rich and varied. These foundations include the study of algebras and algebraic structures associated with rough sets (Banerjee and Chakraborty 2004; Cattaneo and Ciucci 2004), entropy (Bianucci et al. 2007; Ślęzak 2002), lattice theory (Chakraborty and Banerjee 2007; Järvinen 2007), similarity and indiscernibility (Orłowska 1985b), rough-satisfiability (Gomolinska 2005), logic (Düntsch 1997; Polkowski 2002), and rough mereology (Polkowski and Skowron 1996). In general, an overview of the mathematical foundations of rough sets is given by Lech Polkowski (Polkowski 2002; Polkowski and Skowron 1996).

In addition to providing a fairly comprehensive inventory of some of the highlights of rough set-based research, this chapter gives a capsule view of several of the hallmarks of rough set theory, namely, approximation, approximation spaces, discovery of rough sets, and feature set reduction (reducts). For many other issues related to rough set theory and its applications, the reader is referred to the bibliography on rough sets.

This chapter has the following organization. The rough set approach to approximation spaces is given in section “Approximation Spaces”. This

is followed in section “[Rough Sets](#)” by an overview of the general notion of rough sets. The rough set approach to dimensionality reduction is briefly presented in section “[Dimensionality Reduction](#)”. A capsule view of the future of rough sets and their applications is given in section “[Future Directions](#)”.

Approximation Spaces

In rough set theory, approximations are carried out within the context of an approximation space $(O, \mathcal{F}, \sim_B)$, where O is a set of objects, $\mathcal{F}$ is a set of functions representing object features, and $\sim_B$ is an indiscernibility relation defined relative to $B \subseteq \mathcal{F}$. This space is considered fundamental because it provided a framework for the original rough set theory (Pawlak 1981). It has also been observed that an approximation space is the formal counterpart of perception (Orłowska 1985a). Approximation starts with the partition $O/\sim_B$ of O defined by the relation $\sim_B$. Next, any set $X \subseteq O$ is approximated by considering the relation between X and the classes $[x]_B \in O/\sim_B, x \in O$.

An approximation space $(O, \mathcal{F}, \sim_B)$ defined relative to a set of perceptual objects O , a set of functions $\mathcal{F}$, and relation $\sim_B$, has the following basic framework that facilitates observations concerning sample objects.

Framework for an Approximation Space

- O = Set of perceived objects,
- $\mathcal{F}$ = Set of probe functions objects,
- $B \subseteq \mathcal{F}$,
- $\sim_B = \{(x, x') \in X \times X \mid f(x) = f(x') \text{ for any } f \in B\}$,
- $O/\sim_B = \{[x]_B \mid x \in O, B \subseteq \mathcal{F}, B - \text{partition of } O,$
 $[x]_B \in O/\sim_B,$
- $X \subseteq O$, set of objects of interest,
- $B_*X = \bigcup_{x:[x]_B \subseteq X} [x]_B, B - \text{upper approximation},$
- $B^*X = \bigcup_{x:[x]_B \cap X \neq \emptyset} [x]_B, B - \text{upper approximation},$
- $\text{Bnd}_B X = B^*X \setminus B_*X, B - \text{boundary region}.$

Affinities between objects of interest in the set $X \subseteq O$ and classes in the quotient set $O/\sim_B$ can be discovered by identifying those classes that have objects in common with X . Approximation of the set X begins by determining which elementary sets $[x]_B \in O/\sim_B$ are subsets of X . This discovery process leads to the B -lower approximation of $X \subseteq O$ denoted by B_*X . The B -upper approximation B^*X is a sum of equivalence classes $[x]_B \in O/\sim_B$, where each class included in B^*X contains at least one object with a description that matches the description of an object in X . The lower and upper approximations of X provide a basis for defining the boundary of an approximation. Notice that B_*X is always a subset of B^*X .

Several generalizations of the classical rough set approach based on approximation spaces defined as pairs of the form $(O, \mathcal{F}, \sim_B), B \subseteq \mathcal{F}$, have been reported in the literature (Peters et al. 2007; Polkowski 2002; Skowron and Stepaniuk 1996). Among them it is worthwhile mentioning the rough set approach based on similarity (tolerance) relations and the approach to approximation of vague concepts based on the adaptive extension of approximation spaces (Skowron and Stepaniuk 1996) and approximations spaces that consider the nearness of objects (Peters et al. 2007).

Moreover, the approach based on inclusion functions has been generalized to the *rough mereological approach* (Pal et al. 2004; Polkowski 2002; Polkowski and Skowron 1996). The inclusion relation $x\mu_r y$ with the intended meaning *x is a part of y to a degree at least r* has been taken as the basic notion of the rough mereology being a generalization of the Leśniewski mereology (Leśniewski 1929). Research on rough mereology has shown importance of another notion, namely *closeness* of compound objects (e. g., concepts). This can be defined by $x\text{cl}_{r,r'} y$ if and only if $x\mu_r y$ and $y\mu_{r'} x$.

Rough mereology offers a methodology for synthesis and analysis of objects in a distributed environment of intelligent agents, in particular, for synthesis of objects satisfying a given specification to a satisfactory degree or for control in such a complex environment. Moreover, rough

mereology has been used for developing the foundations of the *information granule calculi*, aiming at formalization of the Computing with Words paradigm, recently formulated by Lotfi Zadeh (2001, 2005). More complex information granules are defined recursively using already defined information granules and their measures of inclusion and closeness. Information granules can have complex structures like classifiers or approximation spaces. Computations on information granules lead to the discovery of relevant information granules, e. g., information granules useful in compound concept approximations, pattern recognition and in machine learning. For example, families of approximation spaces labeled by some parameters are considered. By tuning such parameters according to chosen criteria (e. g., minimal description length), one can search for an optimal approximation space.

Rough Sets

In its original conception by Zdzisław Pawlak during the early 1980s, the discovery of the rough set approach to classification of sets of objects resulted from work on knowledge description systems begun by Pawlak during the early 1970s. In the rough set approach, the roughness of our knowledge of a set of objects is gauged in terms of a new approach to approximation that pivots on object descriptions commonly found in information tables also introduced by Pawlak. Approximation is carried out at the level of sets relative to our knowledge about objects of interest. That is, one approximates one set by another set. This approximation is built on a feature space for individual objects. A feature space for objects is defined relative to functions that map sample objects(pre-images) to measurements (images) that gauge the extent of our knowledge about the appearance of the objects. The basic idea is to select a set X of objects of interest and approximate it with another set that resembles X . This means that approximation is carried out at two different levels in the rough set approach.

1. **Object Level.** For a given set of objects design a set of functions $\phi = \{\phi_1, \dots, \phi_n\}$ that represent features of sample objects of interest. The description of an object x is approximated by an n -tuple $(\phi_1(x), \dots, \phi_n(x))$ of feature measurements.
2. **Set Level.** Define the partition of objects in the sample space O using the indiscernibility relation $\sim_B$, where $B \subseteq \phi$. Approximate a set X containing objects of interest relative to the elementary sets in $X/\sim_B$. Approximation of a set X containing objects with features represented by functions ϕ is carried out in terms of the lower approximation B_*X and the upper approximation B^*X .

Example 1 (Sample Approximation of a Set of Coral Pores) For this example, consider the set of coral pores in the coral fragment shown in Fig. 1(b). Then sets $O, X, \mathcal{B}$ and an approximation of X are defined in the following way:

$$\begin{aligned} O &= \{p \mid p \text{ equals a labeled coral pore in Fig.1}\} \\ &= \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}\}, \\ \mathcal{F} &= \{f\}, \text{ where } f: O \rightarrow \mathfrak{R}, \\ f(p_i) &= \text{depth (in mm) of a coral pore } p_i \in X, \\ O/\sim_B &= \{[p_1]_f, [p_4]_f, [p_6]_f\} \text{ shown in Fig.1a,} \\ X &= \{p \mid p \text{ equals a labeled coral pore in Fig.1b}\}, \\ &= \{p_3, p_4, p_5, p_{10}\}. \end{aligned}$$

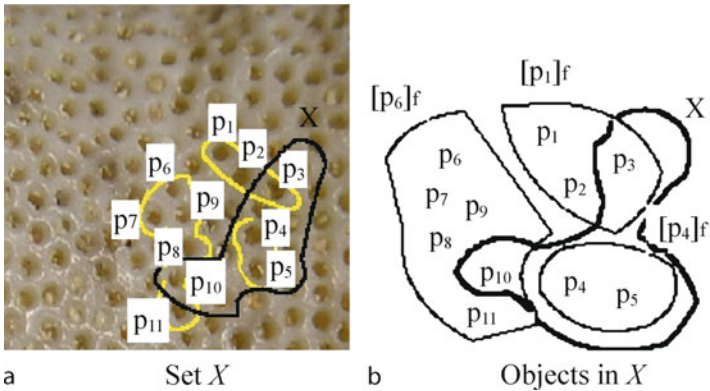
Hence

$$\begin{aligned} B_*X &= [p_4]_f = \{p_4, p_5\}, \\ B^*X &= [p_1]_f \cup [p_4]_f \cup [p_6]_f, \\ \text{Bnd}_B X &= [p_1]_f \cup [p_6]_f. \end{aligned}$$

Then observe that only $[p_4]_f$ is a proper subset of X . In effect, only the class $[p_4]_f$ has a complete affinity with X , since all of the objects $[p_4]_f$ have descriptions that match the descriptions of objects in X . By contrast, classes $[p_1]_f, [p_6]_f$ have only a partial affinity with X , since there are objects in each class that have descriptions that do match the description of any object in X . Observe that the boundary $\text{Bnd}_B X$ is not empty. Hence, X is an example of a rough set.

Rough Sets: Foundations and Perspectives,

Fig. 1 Sample $X \subseteq \mathcal{O}$ to be approximated



Information Tables

For computational reasons, a syntactic representation of information systems is usually given in the form of tables. An information system is represented by the pair $(O, \mathcal{F})$, where O denotes a set of objects (a universe) and $\mathcal{F}$ denotes a set of functions representing features of objects in O . Let $B \subseteq \mathcal{F}$. Discovering objects in the composition of a class $[x]_B, x \in O$ in the partition $O/\sim_B$ in the system $(O, \mathcal{F})$ is accomplished by gathering together inside the class all of those objects that have matching function values. Identifying the classes in $X/\sim_B$ is greatly aided by a table representation of $(O, \mathcal{F})$. This is illustrated with the following example.

Example 2 (Sample Information Table) For this example, assume again that O is a set of words. For simplicity, we limit O to words from poems by John Keats (1795–1821). Consider, for example, the words in the following lines in Part 3 of Keats’ ode *To Autumn* (Keats 1820).

TO AUTUMN
3.
Where are the songs of spring? Aye, where are
they?
Think not of them, thou has thy music too,—
...
John Keats
19 September 1819

By way of approximation of a set of conceptual objects, consider $X \subseteq O$ defined as

$$\begin{aligned} O &= \{x \mid x \text{ is a word in Keats' poetry}\}, \\ &= \{Where, are, the, songs, of, spring, \dots\}, \\ \mathcal{A} &= \{f_1, f_2\}, \end{aligned}$$

where

$$\begin{aligned} f_1(x) &= \begin{cases} 1, & x \text{ begins with a vowel,} \\ 0, & \text{otherwise.} \end{cases} \\ f_2(x) &= \begin{cases} 1, & x \text{ contains two or more vowels,} \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The objects in this example are conceptualized as poetic words and the functions appropriately represent attributes of poetic words. One might say that a word gains in poesy by beginning with a vowel sound (taking away the leading vowel sound tends to undermine the poetic character of a word). Similarly, a repetition of vowel sounds in a word is characteristic of poetic words, especially in the poetry of John Keats. For simplicity, consider only the vowels *a, e, i, o, u*, the vowel sound *ha* in the word *has* and omit consideration of other vowel sounds. In addition, O is restricted to words in Keats’ ode *To Autumn*, partially represented in this Example. The function values for each of the words in the fragment from Keat’s ode are summarized in Table 1. Let $B = \mathcal{A}$. This leads to the following partition of O :

Rough Sets: Foundations and Perspectives,
Table 1 Information System for Keats' Words

O	f_1	f_2	d
Where	0	1	1
Are	1	1	1
The	0	0	0
Songs	0	0	1
Of	1	0	1
Spring	0	0	1
Aye	1	1	1
They	0	0	1
Think	0	0	1
Not	0	0	1
Them	0	0	1
Thou	0	1	1
Has	0	0	0
Thy	0	0	1
Music	0	1	1
Too	0	1	1

$$\begin{aligned}
 [where]_B &= \{where, thou, music, too\} \\
 [are]_B &= \{areAye\} \\
 [the]_B &= \{the, songs, spring, they, think, \\
 &\quad not, them, has, thy\} \\
 [of]_B &= \{of\}
 \end{aligned}$$

Now select, for example,

$$X = \{not, are, Aye, has, of, too\}.$$

This choice of X leads to

$$\begin{aligned}
 B_*X &= [are]_B \cup [of]_B = \{are, Aye, of\} \\
 B^*X &= [are]_B \cup [of]_B \cup [the]_B.
 \end{aligned}$$

In effect, the words in $[are]_B \cup [of]_B$ have an affinity to the words in the set X . In particular, notice that each of the words in the class $[where]_B$

$$[where]_B = \{where, thou, music, too\},$$

have exactly the same f_1, f_2 function values. Similarly, for the composition of the remaining three classes in the partition $X/\sim_B$. Again, for example, notice that the words in class $[the]_B$ have the same function values, namely, $f_1(x) = f_2(x) = 0$, where

x is a word in $[the]_B$. It is a fairly straightforward task to identify the classes extractable from small tables. For large tables, it is necessary to mechanize the class extraction process.

Decision Systems and Decision Rules

Of particular interest is the extension of information systems made possible by including a function d representing what is known as a decision attribute in rough set theory. A decision system is represented by the triple $(O, \mathcal{F}, d)$, where O denotes a set of objects (a universe), $\mathcal{F}$ denotes a set of functions representing features of objects in O , and $d : O \rightarrow V_d$, where V_d is a set of values representing decisions. Presenting decision procedures in a tabular form goes back at least to ancient Babylon. Tabular forms for computer programming dates back to the late 1950s. Next, tabular forms became popular in databases (see <http://www.catalyst.com/products/logicgem/overview.html>). It is clear from the common meaning of the verb *decide* that a decision represents a resolution or determination about something and presumes something knowable by the mind but not necessarily observable by the senses. In more general setting, one can consider a set of decisions D instead of a single decision d .

Example 3 (Sample Decision System) For example, consider an extension of the information system $(O, \mathcal{A})$ with a decision d for Keats' words in Example 3. Recall that O denotes a set of words and $\mathcal{F}$ denotes a set of features representing attributes of words in O . By way of illustration, let d be defined as follows.

$$d(x) = \begin{cases} 1, & x \in O \mid x \text{ is a part of an alliteration,} \\ 0, & \text{otherwise,} \end{cases}$$

where *alliteration* means the occurrence of the same letter or sound at the beginning of adjacent or closely connected words. *Alliteration* is also an example of a concept (i. e., an idea for a class of objects), that provides a basis for decision-making about various words. The common sense view of a concept is the one favored in this introduction to rough sets. Considering *alliteration* to be a concept is consonant with the general meaning of a

theoretical concept (Audi 1999). That is, a *theoretical concept* is a concept expressed by a theoretical term associated with an axiomatic system underlying logic, mathematics, terms drawn from natural language or theories or that constitute the vocabulary of a particular theory such as poetics or a branch of science such as physics. at odds with the notion of concept in logic, mathematics, scientific theory. Table 1 is a sample decision table.

For example, consider the first line of Part 3 of Keats' *To Autumn*:

Where are the songs of spring? Aye, where are they?

The repetition of the $\underline{s}$ sound in *songs* and *Spring* is alliterative. Then $d(\text{songs}) = d(\text{Spring}) = 1$. The word the in the first question in *To Autumn* is not part of an alliteration in that question. Hence, $d(\text{the}) = 0$.

Let $V = \cup\{V_a \mid a \in C\} \cup \{V_d\}$. Atomic formulae over $B \subseteq C \cup D$ and V are expressions $a = v$ called *descriptors (selectors)* over B and V , where $a \in B$ and $v \in V_a$. The set $\mathcal{F}(B, V)$ of formulae over B and V is the least set containing all atomic formulae over B and V and closed with respect to the propositional connectives $\wedge$ (conjunction), $\vee$ (disjunction) and $\neg$ (negation).

By $\|\varphi\|_{\mathcal{A}}$, we denote the meaning of $\varphi \in \mathcal{F}(B, V)$ in the decision table $\mathcal{A}$ which is the set of all objects in O with the property φ . These sets are defined by $\|a = v\|_{\mathcal{A}} = \{x \in O \mid a(x) = v\}$, $\|\varphi \wedge \varphi'\|_{\mathcal{A}} = \|\varphi\|_{\mathcal{A}} \cap \|\varphi'\|_{\mathcal{A}}$; $\|\varphi \vee \varphi'\|_{\mathcal{A}} =$ The formulae from $\mathcal{F}(C, V)$, $\mathcal{F}(d, V)$ are called *condition formulae of $\mathcal{A}$* and *decision formulae of $\mathcal{A}$* , respectively.

Any object $x \in O$ belongs to the *decision class* $\|\wedge_{d \in D} d = d(x)\|_{\mathcal{A}}$ of $\mathcal{A}$. All decision classes of $\mathcal{A}$ create a partition $O/\mathcal{D}$ of the universe O .

A *decision rule* for $\mathcal{A}$ is any expression of the form $\varphi \Rightarrow \psi$, where $\varphi \in \mathcal{F}(C, V)$, $\psi \in \mathcal{F}(D, V)$, and $\|\varphi\|_{\mathcal{A}} \neq \emptyset$. Formulae φ and ψ are referred to as the *predecessor* and the *successor* of decision rule $\varphi \Rightarrow \psi$. Decision rules are often called "IF...THEN..." rules. Such rules are used in machine learning (see, e. g., Friedman et al. 2001).

Decision rule $\varphi \Rightarrow \psi$ is *true* in $\mathcal{A}$ if and only if $\|\varphi\|_{\mathcal{A}} \subseteq \|\psi\|_{\mathcal{A}}$. Otherwise, one can measure its

truth degree by introducing some inclusion measure of $\|\varphi\|_{\mathcal{A}}$ in $\|\psi\|_{\mathcal{A}}$.

Given two unary predicate formulae $\alpha(x), \beta(x)$ where x runs over a finite set O , Łukasiewicz (Łukasiewicz 1913) proposes to assign to $\alpha(x)$ the value $\text{card}(\|\alpha(x)\|)/\text{card}(O)$, where $\|\alpha(x)\| = \{x \in O : x \text{ satisfies } \alpha\}$. The fractional value assigned to the implication $\alpha(x) \Rightarrow \beta(x)$ is then $\text{card}(\|\alpha(x) \wedge \beta(x)\|)/\text{card}(\|\alpha(x)\|)$ under the assumption that $\|\alpha(x)\| \neq \emptyset$. Proposed by Łukasiewicz, that fractional part was much later adapted by machine learning and data mining literature.

Each object x of a decision system determines a *decision rule*

$$\bigwedge_{a \in C} a = a(x) \Rightarrow \bigwedge_{d \in D} d = d(x). \quad (1)$$

For any decision table $\mathcal{A} = (O, C, d)$ one can consider a *generalized decision function* $\partial_{\mathcal{A}} : O \rightarrow V_d$ defined by

$$\partial_{\mathcal{A}}(x) = \{i : \exists x' \in O[(x', x) \in I(\mathcal{A}) \text{ and } d(x') = i]\}. \quad (2)$$

$\mathcal{A}$ is called *consistent (deterministic)*, if $\text{card}(\partial_{\mathcal{A}}(x)) = 1$, for any $x \in O$. Otherwise $\mathcal{A}$ is said to be *inconsistent (non-deterministic)*. Hence, a decision table is inconsistent if it consists of some objects with different decisions but indiscernible with respect to condition attributes. Any set consisting of all objects with the same generalized decision value is called a *generalized decision class*. Now, one can consider certain (possible) rules (see, e. g. Grzymala-Busse 1992) for decision classes defined by the lower (upper) approximations of such generalized decision classes of $\mathcal{A}$. This approach can be extend, using the relationships of rough sets with the Dempster–Shafer theory (see, e. g., Skowron and Grzymala-Busse 1994), by considering rules relative to decision classes defined by the lower approximations of unions of decision classes of $\mathcal{A}$.

Numerous methods have been developed for different decision rule generation that the reader can find in the literature on rough sets. Usually, one is searching for decision rules (semi) optimal

with respect to some optimization criteria describing quality of decision rules in concept approximations.

In the case of searching for concept approximation in an extension of a given universe of objects (sample), the following steps are typical. When a set of rules has been induced from a decision table containing a set of training examples, they can be inspected to see if they reveal any novel relationships between attributes that are worth pursuing for further research. Furthermore, the rules can be applied to a set of unseen cases in order to estimate their classificatory power.

Dimensionality Reduction

We often face the question whether one or more attributes can be removed and still preserve the basic properties of an information system. Let us express this idea more precisely.

Let $C, D \subseteq A$, be sets of condition and decision attributes respectively. We will say that $C' \subseteq C$ is a D -reduct (reduct with respect to D) of C , if C' is a minimal subset of C such that

$$\gamma(C, D) = \gamma(C', D). \quad (3)$$

The intersection of all D -reducts is called a D -core (core with respect to D). Because the core is the intersection of all reducts, it is included in every reduct. Thus, in a sense, the core is the most important subset of attributes, since none of its elements can be removed without affecting the classification power of attributes. Certainly, the geometry of reducts can be more compound. For example, the core can be empty but there can exist a partition of reducts into a few sets with non empty intersection.

Many other kinds of reducts and their approximations are discussed in the literature (see, e. g., Nguyen 2006; Ślęzak 1996). For example, if one change the condition (3) to $\partial_A(x) = \partial_B(x)$, then the defined reducts are preserving the generalized decision. Other kinds of reducts are preserving, e. g.,: (i) the distance between attribute value vectors for any two objects, if this distance is greater than a given threshold, (ii) the distance

between entropy distributions between any two objects, if this distance exceeds a given threshold (Ślęzak 1996), or (iii) the so called reducts relative to object used for generation of decision rules (Bazan 1998). There are some relationships between different kinds of reducts. If B is a reduct preserving the generalized decision, than in B is included a reduct preserving the positive region. For mentioned above reducts based on distances and thresholds one can find analogous dependency between reducts relative to different thresholds. By choosing different kinds of reducts we select different degrees to which information encoded in data is preserved. Reducts are used for building data models. Choosing a particular reduct or a set of reducts has impact on the model size as well as on its quality in describing a given data set. The model size together with the model quality are two basic components tuned in selecting relevant data models. This is known as the minimal length principle (see, e. g., Rissanen 1985). Selection of relevant kinds of reducts is an important step in building data models. It turns out that the different kinds of reducts can be efficiently computed using heuristics based, e. g., on the Boolean reasoning approach (Nguyen 2006; Skowron and Rauszer 1992).

Summary

We are now observing a growing research interest in the foundations of rough sets that include various logical, algebraic, philosophical complexity issues of rough sets. Some relationships have already been established between rough sets and other approaches as well as a wide range of hybrid systems have been developed.

As a result, rough sets are linked with decision system modeling and analysis of complex systems, fuzzy sets, neural networks, evolutionary computing, data mining and knowledge discovery, pattern recognition, machine learning, data mining, and approximate reasoning, multicriteria decision making. In particular, rough sets are used in probabilistic reasoning, granular computing (including information granule calculi based on rough mereology), intelligent control, intelligent

agent modeling, identification of autonomous systems, and process specification.

A wide range of applications of methods based on rough set theory alone or in combination with other approaches have been discovered in the following areas: acoustics, bioinformatics, business and finance, chemistry, computer engineering (e. g., data compression, digital image processing, digital signal processing, parallel and distributed computer systems, sensor fusion, fractal engineering), decision analysis and systems, economics, electrical engineering (e. g., control, signal analysis, power systems), environmental studies, digital image processing, informatics, medicine, molecular biology, musicology, neurology, robotics, social science, software engineering, spatial visualization, Web engineering, and Web mining.

For further readings on rough set theory and applications the reader is referred to (Pawlak and Skowron 2007; Polkowski 2002) and to books, special issues of journals, issues of Transactions on Rough Sets, and proceedings cited in the bibliography of this article.

Future Directions

One of the main future directions for research based on rough sets in combination with other approaches such as granular computing and other computational intelligence approaches for developing intelligent systems can be found in the framework of Wisdom Technology (Jankowski and Skowron 2007, 2008), for complex vague concept approximation and approximate reasoning about such concepts by agents or teams of agents searching for solutions of problems in real-life dynamically changing (distributed) environments in which these agents are operating. Such systems consist of autonomous agents operating in highly unpredictable environments and interacting with each other.

In addition, a number of new rough set-based research areas have emerged during the past several years such as various forms of learning, including hierarchical learning or analogy based reasoning (Bazan 2008; Bazan et al. 2006b;

Nguyen et al. 2004; Peters et al. 2006; Skowron et al. 2005; Wojna 2007), concept approximation (Bazan et al. 2006a), multicriteria decision making (Greco et al. 2001), entropy in information systems (Bianucci et al. 2007; Ślęzak 2002), rough neural computing (Pal et al. 2004), rough granular computing (Skowron and Peters 2008), and a near set approach to image processing (Henry and Peters 2007; Lockery and Peters 2007).

Acknowledgments The research has been supported by the grant from Ministry of Scientific Research and Information Technology of the Republic of Poland and by grant 185986 from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Bibliography

Primary Literature

- Audi R (1999) The Cambridge dictionary of philosophy, 2nd edn. Cambridge University Press, Cambridge
- Banerjee M, Chakraborty MK (2004) In: Pal SK, Polkowski L, Skowron A (eds) Rough-neural computing. Techniques for computing with words. Springer, Berlin, pp 157–184
- Bazan J (1998) In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery 1: methodology and applications, Studies in fuzziness and soft computing, vol 18. Physica, Heidelberg, pp 321–365
- Bazan J (2008) In: Pedrycz W, Skowron A, Kreinovich V (eds) Handbook of granular computing. Wiley, New York (in press), pp 777–799
- Bazan J, Skowron A, Swiniarski R (2006a) Rough sets and vague concept approximation: from sample approximation to adaptive learning. Transactions on rough sets V, Lecture notes in computer science, vol 4100. Springer, Heidelberg, pp 39–62
- Bazan J, Kruczek P, Bazan-Socha S, Skowron A, Pietrzyk JJ (2006b) In: Greco S, Hato Y, Hirano S, Inuiguchi M, Miyamoto S, Nguyen HS, Słowiński R (eds) Proceedings of the 5th international conference on rough sets and current trends in computing, RSCTC 2006, Kobe, Japan, 6–8 November, Lecture notes in artificial intelligence, vol 4259. Springer, Heidelberg, pp 418–427
- Bianucci D, Cattaneo G, Ciucci D (2007) Fund Inform 75: 77–105
- Cattaneo G, Ciucci D (2004) Algebraic structures for rough sets. Transactions on rough sets II. LNCS, Lecture notes in artificial intelligence, vol 3135. Springer, Heidelberg, pp 63–101
- Chakraborty MK, Banerjee M (2007) Fund Inform 75: 123–139

- Devlin J (1993) The joy of sets. Fundamentals of contemporary set theory. Springer, Berlin
- Dütsch I (1997) *Theor Comput Sci* 179:427–436
- Frege G (1879) *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Nebert, Halle
- Friedman JH, Hastie T, Tibshirani R (2001) The elements of statistical learning: data mining, inference, and prediction. Springer, Heidelberg
- Gomolinska A (2005) Rough validity, confidence, and coverage of rules in approximation spaces. *Transactions on rough sets III*. LNCS, Lecture notes in artificial intelligence, vol 3400. Springer, Heidelberg, pp 57–81
- Greco S, Matarazzo B, Słowiński R (2001) *Eur J Oper Res* 129:1–47
- Greco S, Słowiński R, Stefanowski J, Zurawski M (2005) Incremental vs. non-incremental rule induction for multicriteria classification. *Transactions on rough sets II*. LNCS, Lecture notes in artificial intelligence, vol 3135. Springer, Heidelberg, pp 33–53
- Grzymala-Busse JW (1992) In: Słowiński R (ed) *Intelligent decision support – handbook of applications and advances of the rough sets theory, system theory, knowledge engineering and problem solving*, vol 11. Kluwer, Dordrecht, pp 3–18
- Grzymala-Busse JW, Grzymala-Busse WJ (2007) An experimental comparison of three rough set approaches to missing attribute values. *Transactions on rough sets VI*. LNCS, Lecture notes in artificial intelligence, vol 4374. Springer, Heidelberg, pp 31–50
- Henry C, Peters JF (2007) Proceedings of eleventh international conference on rough sets, fuzzy sets, data mining and granular computing. Joint rough set symposium, Lecture notes in artificial intelligence, vol 4482. Springer, Berlin, pp 475–482
- Int. Rough set society website: www.roughsets.org
- Jankowski A, Skowron A (2007) In: Ehrenfeucht A, Marek V, Srebrny M (eds) *Andrzej Mostowski and foundational studies*. IOS Press, Amsterdam, pp 106–143
- Jankowski A, Skowron A (2008) In: Pedrycz W, Skowron A, Kreinovich V (eds) *Handbook of granular computing*. Wiley, New York, pp 329–345
- Järvinen J (2007) Lattice theory for rough sets. *Transactions on rough sets VI*. LNCS, Lecture notes in artificial intelligence, vol 4374. Springer, Heidelberg, pp 400–496
- Keats J (1820) *Lamia, Isabella, the eve of St. Agness, and other poems*. Taylor and Hessey, London
- Konrad E, Orlowska E, Pawlak Z (1981) Knowledge representation systems. Definability of informations. Prace IPI Pan, ICS PAS report, vol 433 April. Institute of Computer Science, Polish Academy of Sciences, Warsaw
- Leśniewski S (1929) *Fundam Math* 14:1–81
- Lockery D, Peters JF (2007) Proceedings of eleventh international conference on rough sets, fuzzy sets, data mining and granular computing. Joint rough set symposium, Lecture notes in artificial intelligence, vol 4482. Springer, Berlin, pp 483–490
- Łukasiewicz J (1913) In: Borkowski L (ed) *Jan Łukasiewicz – selected works*. North Holland, Amsterdam; (1970). Polish Scientific Publishers, Warsaw, pp 16–63
- Murray JA, Bradley H, Craigie W, Onions C (1933) *The Oxford English dictionary*. Oxford University Press, London
- Nguyen HS (2006) Approximate boolean reasoning: foundations and applications in data mining. *Transactions on rough sets V*. LNCS, Lecture notes in artificial intelligence, vol 4100. Springer, Heidelberg, pp 344–523
- Nguyen SH, Bazan J, Skowron A, Nguyen HS (2004) Layered learning for concept synthesis, *transactions on rough sets I*. LNCS, Lecture notes in artificial intelligence, vol 3100. Springer, Heidelberg, pp 187–208
- Orłowska E (1985a) In: Dorn G, Weingartner P (eds) *Foundations of logic and linguistics. Problems and solutions*. Plenum, London, pp 465–482
- Orłowska E (1985b) A logic of indiscernibility relations, *Lecture notes in computer science*, vol 208. Springer, Berlin, pp 177–186
- Pal SK, Polkowski L, Skowron A (2004) *Rough-neural computing. Techniques for computing with words*. Springer, Berlin
- Pawlak Z (1973) Mathematical foundations of information retrieval. In: *Proceedings of symposium of mathematical foundations of computer science*, 3–8 September, High Tartras, pp. 135–136. See also: *Mathematical foundations of information retrieval*. Computation Center, Polish Academy of Sciences, Research Report CC PAS Report 101, Warsaw
- Pawlak Z (1981) Classification of objects by means of attributes. Institute for Computer Science, Polish Academy of Sciences. Research report CC PAS report 429, Warsaw
- Pawlak Z (1991) *Rough sets. Theoretical aspects of reasoning about data*. Kluwer, Dordrecht
- Pawlak Z (1995) On rough derivatives, rough integrals, and rough differential equations, *ICS research report* 41/95. Institute of Computer Science, Warsaw
- Pawlak Z, Skowron A (2007) *Inf Sci Int J* 177(1):41–73
- Peters JF (2008) *Int J Inf Technol Intell Comput* 3(2):1–35
- Peters JF, Henry C (2007) *Int J Intell Real-Time Autom* 20(5):667–675
- Peters JF, Henry C, Gunderson DS (2006) *Int J Hybrid Intell Syst* 3:1–14
- Peters JF, Skowron A, Stepaniuk J (2007) *Fund Inform* 79(3–4):497–512
- Polkowski L (2002) *Rough sets: mathematical foundations*. Advances in soft computing. Physica, Heidelberg
- Polkowski L, Skowron A (1996) *J Approx Reason* 15(4): 333–365
- Polkowski L, Skowron A (1998a) *Rough sets in knowledge discovery 1: methodology and applications*. Studies in fuzziness and soft computing, vol 18. Physica, Heidelberg

- Polkowski L, Skowron A (1998b) Rough sets in knowledge discovery 2: applications, case studies and software systems. Studies in fuzziness and soft computing, vol 19. Physica, Heidelberg
- Rissanen J (1985) In: Kotz S, Johnson N (eds) Encyclopedia of statistical sciences. Wiley, New York, pp 523–527
- Rough Set Database: <http://rds.univ.rzeszow.pl/>
- Skowron A, Grzymala-Busse JW (1994) In: Yager R, Fedrizzi M, Kacprzyk J (eds) Advances in the Dempster-Shafer theory of evidence. Wiley, New York, pp 193–236
- Skowron A, Peters JF (2008) In: Pedrycz W, Skowron A, Kreinovich V (eds) Handbook of granular computing. Wiley, pp 285–328
- Skowron A, Rauszer C (1992) In: Słowiński R (ed) Intelligent decision support – handbook of applications and advances of the rough sets theory, system theory, knowledge engineering and problem solving, vol 11. Kluwer, Dordrecht, pp 331–362
- Skowron A, Stepaniuk J (1996) Fund Inform 27:245–253
- Skowron A, Swiniarski R, Synak P (2005) Approximation spaces and information granulation. Transactions on rough sets III. LNCS, Lecture notes in artificial intelligence, vol 3400. Springer, Heidelberg, pp 175–189
- Skowron A, Stepaniuk J, Peters JF, Swiniarski R (2006) Fund Inform 72(1–3):363–378
- Ślęzak D (1996) Sixth international conference on information processing and management of uncertainty in knowledge-based Systems, IPMU'1996, vol III. Granada, Spain, pp 1159–1164
- Ślęzak D (2002) Fund Inform 53:365–387
- Stepaniuk J (1998) In: Polkowski L, Skowron A (eds) Rough sets in knowledge discovery 2: applications, case studies and software systems, studies in fuzziness and soft computing, vol 19. Physica, Heidelberg, pp 109–126
- Stepaniuk J (2000) In: Polkowski L, Tsumoto S, Lin TY (eds) Rough set methods and applications. New developments in knowledge discovery in information systems. Physica, Heidelberg, pp 137–233
- Stepaniuk J (2007) Approximation spaces in multi relational knowledge discovery, transactions on rough sets VI: LNCS, Lecture notes in artificial intelligence, vol 4374. Springer, Heidelberg, pp 351–365
- Transactions on Rough Sets. Website: <http://www.springer.com/east/home/computer/lncs?SGWID=5-164-6-99627-0>, special issues of other journals, proceedings of international conferences, tutorials or surveys (see, references in this paper and, e. g., Pawlak Z, Skowron A (2007) Rudiments of rough sets. Inf Sci Int J 177(1): 3–27; Rough sets: some extensions. Inf Sci Int J 177(1): 28–40; Rough sets and boolean reasoning. Inf Sci Int J 177(1):41–73, and numerous web pages (www.roughsets.org, and logic.mimuw.edu.pl))
- University of Regina Electronic Bulletin of the Rough Set Community: <http://www.cs.uregina.ca/~Eroughset>
- Vinogradov IM (ed) (1995) Encyclopedia of mathematics. Kluwer, Dordrecht
- Warsaw University Logic Group. Website: logic.mimuw.edu.pl
- Wojna A (2007) Analogy-based reasoning in classifier construction. Transactions on rough sets IV. LNCS, Lecture notes in artificial intelligence, vol 3700. Springer, Heidelberg, pp 277–374
- Zadeh LA (2001) AI Mag 22(1):73–84
- Zadeh LA (2005) Inf Sci 171:1–40
- Ziarko W (1993) J Comput Syst Sci 46:39–59

Further Readings – Books

- Demri S, Orłowska E (2002) Incomplete information: structure, inference, complexity. Monographs in theoretical computer science. Springer, Berlin
- Dunin-Kępicz B, Jankowski A, Skowron A, Szczuka M (eds) (2005) Monitoring, security, and rescue tasks in multiagent systems, MSRAS'2004. Advances in soft computing. Springer, Heidelberg
- Dütsch I, Gediga G (2000) Rough set data analysis: a road to non-invasive knowledge discovery. Methodos, Bangor
- Grzymala-Busse JW (1990) Managing uncertainty in expert systems. Kluwer, Norwell
- Inuiguchi M, Hirano S, Tsumoto S (eds) (2003) Rough set theory and granular computing, studies in fuzziness and soft computing, vol 125. Springer, Heidelberg
- Kostek B (2005) Perception-based data processing in acoustics. Applications to music information retrieval and psychophysiology of hearing, studies in computational intelligence, vol 3. Springer, Heidelberg
- Lin TY, Cercone N (eds) (1997) Rough sets and data mining – analysis of imperfect data. Kluwer, Boston
- Lin TY, Yao YY, Zadeh LA (eds) (2001) Rough Sets, Granular Computing and Data Mining. Studies in Fuzziness and Soft Computing. Physica, Heidelberg
- Mitra S, Acharya T (2003) Data mining. Multimedia, soft computing, and bioinformatics. Wiley, New York
- Munakata T (ed) (1998) Fundamentals of the new artificial intelligence: beyond traditional paradigms, Graduate texts in computer science, vol 10. Springer, New York
- Orłowska E (ed) (1997) Incomplete information: rough set analysis. Studies in fuzziness and soft computing, vol 13. Physica, Heidelberg
- Pal SK, Skowron A (eds) (1999) Rough fuzzy hybridization: a new trend in decision-making. Springer, Singapore
- Polkowski L, Lin TY, Tsumoto S (eds) (2000) Rough set methods and applications: new developments in knowledge discovery in information systems, studies in fuzziness and soft computing, vol 56. Physica, Heidelberg
- Słowiński R (ed) (1992) Intelligent decision support – handbook of applications and advances of the rough sets theory, system theory, knowledge engineering and problem solving, vol 11. Kluwer, Dordrecht
- Zhong N, Liu J (eds) (2004) Intelligent Technologies for Information Analysis. Springer, Heidelberg

Further Readings – Transactions on Rough Sets

Peters JF, Skowron A et al (eds) (2004–2008) Transactions on rough sets I–VIII. (Lecture notes in computer science), vols 3100, 3135, 3400, 3700, 4100, 4374, 4400, 5084. Springer, Heidelberg

Further Readings – Special Issues of Journals

Cercone N, Skowron A, Zhong N (eds) (2001) Comput Intell Int J 17(3)
 Lin TY (ed) (1996) J Intell Autom Soft Comput 2(2)
 Pal SK, Pedrycz W, Skowron A, Swiniarski R (eds) (2001) Neurocomputing 36
 Peters J, Skowron A (eds) (2001) Int J Intell Syst 16(1)
 Skowron A, Pal SK (eds) (2003) Special volume: rough sets, pattern recognition and data mining. Pattern Recogn Lett 24(6)
 Słowiński R, Stefanowski J (eds) (1993) Found Comput Decis Sci 18(3–4)
 Ziarko W (ed) (1995) Comput Intell Int J 11(2)
 Ziarko W (ed) (1996) Fundam Inform 27(2–3)

Further Readings – Proceedings of International Conferences

Alpighini JJ, Peters JF, Skowron A, Zhong N (eds) (2002) Third international conference on rough sets and current trends in computing, RSCTC'2002, Malvern, PA, 14–16 October, Lecture notes in artificial intelligence, vol 2475. Springer, Heidelberg
 An A, Stefanowski J, Ramanna S, Butz CJ, Pedrycz W, Wang G (eds) (2007) Proceedings of the eleventh international conference on rough sets, fuzzy sets, data mining, and granular computing, RSFD GrC 2007, Toronto, Canada, 14–16 May, Lecture notes in artificial intelligence, vol 4482. Springer, Heidelberg
 Greco S, Hata Y, Hirano S, Inuiguchi M, Miyamoto S, Nguyen HS, Slowinski R (eds) (2006) Proceedings of the fifth international conference on rough sets and current trends in computing, RSCTC 2006, Kobe, Japan, 6–8 November, Lecture notes in artificial intelligence, vol 4259. Springer, Heidelberg
 Hirano S, Inuiguchi M, Tsumoto S (eds) (2001) Bull Int Rough Set Soc 5(1–2)
 Kryszkiewicz M, Peters JF, Rybinski H, Skowron A (eds) (2007) Proceedings of the international conference on rough sets and intelligent systems paradigms, RSEISP 2007, Warsaw, Poland, 28–30 June, Lecture notes in artificial intelligence, vol 4585. Springer, Heidelberg
 Lin TY, Wildberger AM (eds) (1995) Soft computing: rough sets, fuzzy logic, neural networks, uncertainty management, knowledge discovery. Simulation Councils, USA
 Polkowski L, Skowron A (eds) (1998) First international conference on rough sets and soft computing, RSCTC'1998, Lecture notes in artificial intelligence, vol 1424. Springer, Poland
 Skowron A (ed) (1985) Proceedings of the 5th symposium on computation theory. Zaborów, Poland, 1984, Lecture notes in computer science, vol 208. Springer, Berlin
 Skowron A, Szczuka M (eds) (2003) Electron Notes Comput Sci, 82(4). Elsevier, Netherlands, <http://www.elsevier.nl/locate/entcs/volume82.html>
 Słezak D, Wang G, Szczuka M, Düntsch I, Yao Y (eds) (2005a) Proceedings of the 10th international conference on rough sets, fuzzy sets, data mining, and granular computing, RSFD GrC'2005, Regina, Canada, 31 August–3 September, Part I, Lecture notes in artificial intelligence, vol 3641. Springer, Heidelberg
 Słezak D, Yao JT, Peters JF, Ziarko W, Hu X (eds) (2005b) Proceedings of the 10th international conference on rough sets, fuzzy sets, data mining, and granular computing, RSFD GrC'2005, Regina, Canada, 31 august–3 September, part II, Lecture notes in artificial intelligence, vol 3642. Springer, Heidelberg
 Terano T, Nishida T, Namatame A, Tsumoto S, Ohsawa Y, Washio T (eds) (2001) New Frontiers in artificial intelligence, joint JSAI'2001. Workshop post-proceedings, Lecture notes in artificial intelligence, vol 2253. Springer, Heidelberg
 Tsumoto S, Kobayashi S, Yokomori T, Tanaka H, Nakamura A (eds) (1996) Proceedings of the the fourth internal workshop on rough sets, fuzzy sets and machine discovery, University of Tokyo, Japan, 6–8 November. The University of Tokyo, Tokyo
 Tsumoto S, Słowiński R, Komorowski J, Grzymala-Busse J (eds) (2004) Proceedings of the 4th international conference on rough sets and current trends in computing, RSCTC'2004, Uppsala, Sweden, 1–5 June, Lecture notes in artificial intelligence, vol 3066. Springer, Heidelberg
 Wang G, Liu Q, Yao Y, Skowron A (eds) (2003) Proceedings of the 9th international conference on rough sets, fuzzy sets, data mining, and granular computing, RSFDGrC'2003, Chongqing, China, 26–29 may, Lecture notes in artificial intelligence, vol 2639. Springer, Heidelberg
 Yao JT, Lingras P, Wu WZ, Szczuka M, Cercone NJ, Słezak D (eds) (2007) Proceedings of the second international conference on rough sets and knowledge technology, RSKT 2007, Toronto, Canada, 14–16 may, Lecture notes in artificial intelligence, vol 4481. Springer, Heidelberg
 Zhong N, Skowron A, Ohsuga S (eds) (1999) Proceedings of the 7th international workshop on rough sets, fuzzy sets, data mining, and granular-soft computing, RSFDGrC'99, Yamaguchi, 9–11 November, Lecture notes in artificial intelligence, vol 1711. Springer, Heidelberg
 Ziarko W (ed) (1994) Rough sets, fuzzy sets and knowledge discovery: proceedings of the second international workshop on rough sets and knowledge discovery, RSKD'93, Banff, Canada, 12–15 October 1993. Workshops in computing. Springer & British Computer Society, London/Berlin
 Ziarko W, Yao Y (2001) Proceedings of the 2nd international conference on rough sets and current trends in computing, RSCTC'2000, Banff, Canada, 16–19 October 2000, Lecture notes in artificial intelligence, vol 2005. Springer, Heidelberg

Introduction to Soft Computing

Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

As science progresses, scholars and researchers continue to develop models of various phenomena and problems. These models have become more and more formal, notably mathematical, and hence have relied more and more on computing.

The initial period has been characterized by computing means as well-defined algorithmic procedures implemented by using computing architectures, tools, and techniques based on traditional binary (yes–no or 0–1) logic. This logic has also been a basis for all kinds of approaches within the field of artificial intelligence, a field of science and technology that has been vividly developed since the mid-1900s and aimed at devising machines capable of performing human-specific tasks that require intelligence. Symbolic computation has become a main paradigm within artificial intelligence, and the role of numerical computing has been neglected to a large extent.

As promises of artificial research gurus that machines will be able to perform virtually all sophisticated human-specific tasks and practically replace humans in a short time have failed, it has become more and more evident that sticking to symbolic computations only is a mistake. Clearly, there may be tasks for which symbolic computations are adequate, but there may be tasks, presumably even more, when numerical computations are indispensable.

Even if this necessity of a synergistic use of symbolic and numerical computations in the construction of intelligent systems has become obvious, people have been, for a long time, not fully convinced that, since the human being is a crucial element in all kinds of research, analysis and

construction of broadly perceived intelligent systems, computational techniques, and processes involved therein should reflect human-specific features.

In our context, the main feature of the human element in computation is that natural language is the only fully natural means of articulation and communication of a human being, and natural language is plagued by inherent vagueness, ambiguity, imprecision, etc. These kinds of information imperfections are out of scope of conventional mathematical tools based on – for instance – probabilistic or statistical tools.

The necessity of developing a new framework which would differ from the traditional “hard” computing based on traditional mathematics, i.e., basically on binary logic, and the dichotomy of true and false, has been recognized quite early. However, it was Lotfi A. Zadeh from the University of California at Berkeley who first pronounced that necessity – in the context of the analysis of complex animate systems, notably human centered – in a more comprehensive way in the early 1960s. In 1965, he introduced fuzzy sets theory, presumably the first comprehensible, yet simple and efficient calculus of imprecision. A fuzzy set is a class to which elements can belong to a degree, from full membership, through all intermediate values, to full nonmembership. Clearly, this has provided a constructive means for the representation of imprecisely specified concepts, relations, reasoning schemes, etc. Zadeh then proceeded to the formulation of possibility theory, and finally based his ideas on a more direct link to natural language which has culminated in his computing with words and perceptions. These and closely related subjects constitute the major topics of the volume *Granular, Fuzzy and Soft Computing*, which forms a part of the complexity series.

The introduction of a fuzzy set, followed by numerous applications of fuzzy logic-based tools and techniques, notably for control, ranging from the control of domestic appliances and automatic transmissions in cars to the control of industrial

installations, has certainly been very important for the proliferation of fuzzy logic, maybe even more generally for nonconventional computational tools and techniques.

As a next step, Zadeh has advocated a wider view of those new computational tools and techniques and coined the term of *soft computing*. Basically, soft computing may be viewed as a departure from conventional (hard and “precise”) computing in that it is tolerant of imprecision, uncertainty, partial truth, and approximate relationships, approximately satisfied properties, etc.

The main principle underlying soft computing is to be able to exploit a tolerance for imprecision, uncertainty, partial truth, and approximation to attain tractable, robust, and computationally inexpensive modeling and solution tools and techniques. A natural connection to neural computing and evolutionary computing has more recently been recognized.

It is usually assumed that the principal components of soft computing are fuzzy logic (complemented with rough sets theory which may be viewed to address related problems of imprecise information but understood and dealt with in a different way), neural computation, and evolutionary computation. These areas are sometimes complemented with machine learning and probabilistic reasoning, with the latter subsuming belief networks, chaos theory, and parts of learning theory. Moreover, for a long time, rough sets have been considered to be a relevant part of soft computing.

However, it should be noticed that soft computing is not a confluence of more or less independent fields, but is meant as a synergistic combination of various methodologies and methods in which each one contributes in a synergistic and complementary way to the analysis and solution of a problem.

In this section, we will present the main components of soft computing. The concepts of synergy and complementarity mentioned above will provide a pivotal perspective.

We start with a brief introduction to fuzzy sets and systems (see ► [“Foundations of Fuzzy Sets Theory”](#)). First, we define the concept of a fuzzy set followed by the definitions of its main

properties. The idea of a fuzzy relation is then introduced, and shown to be a convenient apparatus for devising approximate reasoning schemes. The concept of a linguistic variable is shown to be a basis for a natural-language-based approach to modeling and reasoning. Fuzzy arithmetic is presented making it possible to use imprecisely specified (fuzzy) numbers in models and algorithms. A universal problem of decision-making is dealt with under fuzzy goals, constraints, etc., in a static and dynamic context.

Though fuzzy sets in the traditional Zadeh sense do provide a simple yet efficient means for the representation and handling of imprecision, some natural extensions of the basic concept of a fuzzy set have appeared. One of the most relevant and popular is the concept of a type 2 fuzzy set in which the degree of membership itself, which is a real number from the unit interval in the traditional fuzzy set, is now a fuzzy set defined in the unit interval. This natural extension has implied relevant extensions of fuzzy modeling and reasoning tools and techniques which are presented in detail, and potentials for applications are underlined (see ► [“Fuzzy Logic, Type-2 and Uncertainty”](#)).

As in any formal model in which, as is true in virtually all cases in reality, multiple aspects, points of view, and entities have to be accounted for, a proper aggregation is of a paramount importance. A comprehensive account of various aggregation operators is provided, both from a strictly mathematical point of view and from a more constructive, application-oriented one (see ► [“Aggregation Operators and Soft Computing”](#)).

The discussion of the foundations of fuzzy logic is completed with a brief account of possibility theory initiated by Zadeh in the late 1970s as an extension of fuzzy sets theory and fuzzy logic, a mathematical theory for dealing with certain types of imperfect (uncertain) information in a way that is an alternative to probability theory. Main concepts related to possibility theory are presented, and a historical perspective is provided (see ► [“Possibility Theory”](#)).

The comprehensive presentation of foundational concepts and properties related to fuzzy logic is then followed by sections presenting

some more relevant areas which can be of use in the analysis of many systems and in the solution of many problems in diverse areas of science and technology.

First, fuzzy optimization and fuzzy mathematical programming are outlined starting with basic concepts of optimality, feasibility, etc., under fuzzy information. Then, various classes of fuzzy optimization and fuzzy mathematical programming problem classes are briefly outlined and basic solution concepts and algorithms are presented. Some examples of applications are mentioned (see ► [“Fuzzy Optimization”](#)).

The topic of statistics with imprecise data is considered next, a very interesting topic from a conceptual point of view, and practically relevant. First, a sound motivation is presented by pointing out that, in reality, much available data is imprecise, and so they cannot be used by powerful and well-founded, yet too “hard” traditional statistical tools and techniques. New concepts are formulated and extensions of well-known statistical means under an imprecise information are presented. Some applications are outlined (see ► [“Statistics with Imprecise Data”](#)).

The next papers are concerned with the two other key elements of soft computing: neural computation and evolutionary computation. To emphasize their synergy and complementarity, these tools have been presented in a hybrid context, that is, as a presentation of new fields in which the confluence of fuzzy, neural, and evolutionary tools yields a new quality, and results in the emergence of a new class of tools and techniques which have already changed the landscape of soft computing.

We start with the neuro-fuzzy systems, which refer to a combination of (artificial) neural networks and fuzzy logic, and may be viewed as a main, fundamental step into the development of hybrid intelligent systems (see ► [“Neuro-fuzzy Systems”](#)). Basically, neuro-fuzzy systems try to combine transparent, human-like reasoning types of fuzzy rule-based systems with a parallel, connectionist type of neural networks. Therefore, they are universal approximators with an ability to yield interpretable IF-THEN rules, bridging the gap between accuracy and interpretability.

A broad coverage of main architectures and implementations is presented.

Though neuro-fuzzy systems have been a considerable step forward toward in developing modeling tools of a better expressive power, effectiveness, and efficiency, some natural extensions have been proposed that take advantage of the increasing popularity of evolutionary computation. It has been shown theoretically and experimentally that the incorporation in broadly perceived fuzzy systems, including neuron-fuzzy systems, of a learning mechanism which can make it possible to adjust the systems’ structure and parameters may greatly enhance performance. The use of evolutionary computation has led in this context to the concept of an evolving fuzzy system that has been considered a promising alternative for some time (see ► [“Evolving Fuzzy Systems”](#)).

The above mentioned hybridization techniques, which basically boil down to architectures and algorithms being a synergistic combination of fuzzy systems, notably fuzzy rule-based systems, neural networks, and evolutionary computation, in particular evolutionary learning, can be further extended when applied to real-world problems which always provide much inspiration. Applications for systems modeling and control are of a universal importance for many fields of science and technology and hence are dealt with in detail (see ► [“Hybrid Soft Computing Models for Systems Modeling and Control”](#)).

The entries briefly described above have been related to fuzzy sets and some of their hybridizations by including elements of neural networks and evolutionary computation to attain some “superadditivity” of the strengths and power of the particular single tools and techniques.

Rough sets theory, introduced by the late Zdzislaw Pawlak in the early 1980s, has been for some time considered a very promising tool for dealing with imperfect information, notably in data analysis and decision-making.

Basically, a rough set is a formal approximation of a conventional crisp set in that it is represented as a pair of two traditional crisp sets which constitute the lower and the upper approximation of the original. These can be viewed as sets of elements that

possibly and surely belong to the original set. Rough sets theory is presented in much detail, first in a more traditional, pure setting in Pawlak's spirit. Various concepts are presented, relations between them are formulated and proved, and their possible applications are outlined, notably in the relevant areas of data mining and knowledge discovery. Some extensions of the basic concept of a rough set are also presented (see ► [“Rough Sets: Foundations and Perspectives”](#)).

Next, a “meta-problem” in science, that is decision-making, is dealt with in the context of rough sets. In addition to the basic definition of a rough set, which provides much insight into the formalization and solution of various classes of decision-making problems, a novel idea of a dominance-based rough set is presented. Its use to derive new, more human-consistent solution concepts in decision-making and data analysis is proposed (see ► [“Rough Sets in Decision-Making”](#)).

To summarize, we have tried to present the vast and diversified area of soft computing in a comprehensive, illustrative, and constructive way. The

basic perspective assumed may be viewed as presenting soft computing as a significant paradigm shift in computing in that it tries to exploit the fact that the human mind, unlike the computer we have now, has a remarkable ability to store and process information which is predominantly imprecise, uncertain, and granular.

We have started with some basic fundamental concepts and properties, and then have devoted much space to the presentation of what is characteristic for soft computing: a synergistic, complementary use of various tools and techniques to solve problems in an effective and efficient way. This has led to various types of hybridization in which tools from fuzzy systems, rough sets, neural networks, and evolutionary computation have been employed. It seems that soft computing should play a more and more important role in the years to come, in which more and more emphasis should be put on human-consistent and human-centric tools and techniques, and on hybrid systems.

Statistics with Imprecise Data

María Ángeles Gil¹ and Olgierd Hryniewicz²

¹Faculty of Sciences, University of Oviedo, Oviedo, Spain

²Systems Research Institute, Warsaw, Poland

Article Outline

Glossary

Definition of the Subject

Introduction

Mathematical Modeling of Imprecise Data

Fuzzy Random Variables

Statistical Analysis of Fuzzy Data Corresponding to Fuzzy Perceptions of Existing Real-Valued Data

Statistical Analysis of Existing Fuzzy-Valued Data

Future Directions

Bibliography

Glossary

Fuzzy estimators Estimators of “parameters” of probability distributions, or other characteristics, of random variables/fuzzy random variables (such as e. g. the expected value/the fuzzy expected value) when statistical data are imprecise and are described by means of fuzzy sets.

Fuzzy random variable Random element whose observed values are described by fuzzy sets.

Fuzzy set Generalization of a classical notion of a set. In contrast to the case of a classical set, each element x of a fuzzy set may belong to it to a degree described by the so-called membership function $\mu(x)$. Thus, the fuzzy set may be defined as a set of ordered pairs $(x, \mu(x))$, where x belongs to a set $\mathbb{X}$ called the universe of discourse or referential. Alternatively, the

fuzzy set can be identified with its membership function (in the same way that a classical set can be identified with its indicator function).

Fuzzy statistical tests Statistical tests used for the verification of hypotheses about the values of “parameters” of probability distributions, or other characteristics, of random variables/fuzzy random variables when statistical data are imprecise and are described by fuzzy sets.

Fuzzy statistics Generalization of traditional statistics that allows to handle imprecise data described in terms of fuzzy sets.

Imprecise data Data that cannot be described by either real numbers or vectors with real-valued components.

Random sets Random elements whose observed values are sets (e. g. intervals, subsets of a plane).

Definition of the Subject

Traditional statistics deals usually with precisely defined data. The source of these data may be of different nature. Usually the data are collected from observations of random experiments, i. e. experiments whose performance leads to an outcome which cannot be predicted in advance with one hundred percent sureness. When the knowledge about the nature of these random events is available we can use the methods of mathematical statistics and draw conclusions about the source of available data. Uncertainty being intrinsic to random outcomes/events is properly described by using the formalism of the mathematical theory of probability, and generally it is attributed to *future* observations of these outcomes/events.

Traditionally, statistical experimental data are described by real numbers or by vectors whose components are real numbers. These numbers are either observed directly as results of measurements (e. g. height and weight of a person) or correspond to observed counts of certain categories representing labeled events (e. g. a gender of

that person). However, in real life applications the results of measurements are never precise. The precision of every measurement is limited by the precision of a measuring device. In the majority of practical applications this lack of precision may be neglected, and for statistical analysis of available data we can use traditional methods of statistics. If the measurement error cannot be neglected statistical analysis of interval data is recommended by specialists in metrology. However, when statistical data are presented by human beings we need more sophisticated methods for the description of their lack of precision. Therefore, there is a practical need to generalize traditional statistical methods in order to make them applicable for handling imprecise data, e. g. the data described by statements expressed by using a plain language (the so-called “linguistic data”).

Introduction

There is a common agreement that uncertainty characterized by randomness shall be described by considering the theory of probability. Thus, mathematical statistics is a proper tool for dealing with data generated by random experiments and described by precisely defined numbers. However, there also exist other types of uncertainty which are related to vagueness, imprecision, existence of only partial information about experimental outcomes/events of interest, etc. In contrast to randomness, uncertainties of such types are attributed rather to *current* perceptions/observations. It has to be noted that the mathematical modeling of all these types of uncertainty which are different from simple randomness is still the subject of controversies. A good overview of the related problems can be found in the paper by P. Walley (1996). In this article we confine ourselves to the case when randomness is observed together with vagueness understood as the lack of precision.

Specialists in measurement theory recognize different types of uncertainty. For instance, in the ISO/IEC Guide (ISO/IEC 1995) it is recommended to distinguish between uncertainty related to pure randomness and uncertainty of

other nature, such as a lack of precision. However, the lack of precision of statistical data is usually omitted in the statistical analysis of measurements. Consider, for example, the analysis of results of measurements coming from a digital meter. The results of measurements coming from such a meter are always rounded in order to have their representation by a certain number of digits. When we observe a displayed result of a measurement we never know what is the actual value of the measured quantity. What is more important, and often overlooked by statisticians, we may not increase our knowledge about that value by averaging the results of repeated measurements, as it is frequently recommended by statisticians. Consider, for example, a series of measurements showing exactly the same result. Having such results we are in principle not able to distinguish between two fundamentally different cases when the measured quantity is constant and its value belongs to the range of rounding or when it is varying from measurement to measurement with values belonging to that range.

The situation described in the previous paragraph is relatively simple as the range of possible actual values corresponding to the observed result of a measurement is usually precisely defined. There exist, however, situations when either this range is not precisely known or values in the range are not equally compatible with the available data/information/events. We face such cases when results of measurements or classifications are evaluated by humans (e. g., by evaluating indications of an analog meter or classifying individuals in accordance with their height, and so on). The result of a measurement is even more imprecise when it has been obtained without the usage of any meter (e. g., when we visually evaluate the distance between two points); in such situations statistical data may consist of imprecise statements like “around 5 meters”, “more or less between 5 and 10 seconds”, etc. The classification of a person in accordance with his/her height leads often to imprecise assessments, like “very short”, “rather tall”, etc. A similar situation occurs when we deal with retrospective data recalled by human beings; for example, in reliability analysis of field lifetime data we may face situations when failure

times are reported imprecisely using statements like ‘the failure occurred about one month ago’, etc. In all these cases statistical data consist of *imprecise perceptions* of actual real values.

In the previous paragraphs we considered situations when actual values of measured quantities exist, but they are imprecisely perceived. There exist, however, situations when we have to analyze statistical data that represent imprecisely defined concepts. Take for example the color of human hair described using categories such as “blond” or “dark blond”; it is obvious that the border between these two categories is vague; one can try to establish a precise border between these two categories in terms of the results of precise measurements of the spectrum of reflected light, but such attempts seem to be practically senseless. We face a similar situation in classifying a client of a bank office in accordance with his/her degree of aversion to investment which leads to imprecise assessments, like “very low”, “moderate”, etc. In both these cases we could try to collect precise statistical data, e. g., by either asking a respondent to a questionnaire to indicate exactly one choice or coding it numerically. However, it seems to be more prudent, natural and informative to expect imprecise answers to questions pertained to vague notions. If we do so, we may face imprecise statistical data for further analysis. In all these cases statistical data consist of *imprecise actual values* themselves.

There also exists another source of imprecision while dealing with statistical data. For example, in reliability lifetime tests we perform tests in more severe “over-stress” conditions, and then we try to recalculate test results to conditions which are considered “normal”. For this recalculation we can utilize some partial knowledge about possible values of stress-dependent recalculation coefficients. In such a case we have originally precise lifetime data, but after recalculation these data become imprecise.

In all considered cases the lack of precision can be appropriately described by using fuzzy sets introduced by Lotfi A Zadeh. In the second section of the article we recall some basic definitions related to fuzzy sets. We will present the fuzzy sets methodology as a useful tool for the description of

imprecise data. When fuzzy lack of precision is mixed with randomness, either in the sense that available fuzzy data are supposed to come from the perception of real- or vectorial-valued data generated by a random mechanism, or in the sense of they being directly generated by a random mechanism, a convenient tool to use is that of the notion of a fuzzy random variable. This notion is introduced in the third section of this article. The interpretation of the fuzzy random variable depends upon the type of observed fuzzy data and events. In the paper we distinguish between the two types of data which have been described in this section. We start with the description of statistical methods which are useful for the analysis of fuzzy perceptions of existing precise values. Then, we present statistical methods which are useful in the analysis of intrinsically fuzzy-valued data.

Mathematical Modeling of Imprecise Data

There exist competitive methods for the description of vagueness. For example, some statisticians claim that the theory of subjective probability is sufficient for the description of all types of uncertainty. However, many other researchers have shown examples of situations when the application of the classical theory of probability is not sufficient for modeling these situations. Therefore, other formalisms have been proposed for the description of vagueness/imprecision. One of those formalisms, namely the theory of *fuzzy sets* proposed by Lotfi A. Zadeh (1956), has been slowly but widely accepted as a good methodology for the description of imprecise data, both from practical and theoretical (see, e. g., the paper by Terán 2007) points of view.

The basic concept of the theory of fuzzy sets is the universe of discourse or referential $\mathbb{X}$ which may be understood as the set of all possible (feasible) elements that are relevant for the description of a certain concept (quantity). Mathematically speaking a *fuzzy subset* A of a set $\mathbb{X} \neq \emptyset$ (or a fuzzyset, for short) is a map $A : \mathbb{X} \rightarrow [0, 1]$, where $A(x)$ can be interpreted as the degree of

compatibility of x with the ill-defined property characterizing A , or degree of truth of the assertion “ x is A ”, or degree of membership of x to A . Equivalent, but more intuitive for some purposes, a fuzzy set A of $\mathbb{X}$ can be defined as a set of ordered pairs $\{(x, \mu_A)\}$, where $x \in \mathbb{X}$ and $\mu_A : \mathbb{X} \rightarrow [0, 1]$ is the so-called *membership function* of A . In other words, a fuzzy set can be identified with its membership function, in the same way that a classical set can be identified with its indicator function. In what follows, we will consider indistinctly A or μ_A to denote and refer to a fuzzy subset.

Unfortunately, there is no one generally accepted methodology for the construction of membership functions. The majority of researchers assume that the membership function $\mu_A(x)$ is a purely subjective function provided by a person who describes his/her perception of a certain phenomenon or quantity. Some authors provide practical methods for the construction of the membership function when it is interpreted in terms of the theory of possibility as the possibility distribution (see Dubois and Prade 1988 for more information). Some other authors, e.g. Bandemer and Näther (1992) or Viertl (1996), present methods which may be used for the construction of membership functions in a more objective way from measurements of physical quantities. Anyway, our purpose is not entering a discussion here about this point.

In the analysis of imprecise data we are usually interested in the description of interesting phenomena by numbers. For this purpose we can use the concept of a fuzzy number defined as follows (see Dubois and Prade 1978):

Definition 1 The fuzzy subset A of the space of real numbers $\mathbb{R}$, with the membership function $\mu_A : \mathbb{R} \rightarrow [0, 1]$, is a **fuzzy number** if and only if

- (a) A is normal, i. e. there exists at least an element $x_0 \in \mathbb{R}$ such that $\mu_A(x_0) = 1$;
- (b) A is fuzzy convex, i. e., $\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \mu_A(x_1) \wedge \mu_A(x_2)$, for all $x_1, x_2 \in \mathbb{R}$, and $\lambda \in [0, 1]$;
- (c) μ_A is upper semicontinuous;

- (d) the support set, $\text{supp} A = \{x \in \mathbb{R} : \mu_A(x) > 0\}$, is bounded (that is, its closure $\text{cl}(\text{supp} A)$ is compact).

It is easily seen that if A is a fuzzy number then its membership function can be expressed as follows:

$$\mu_A(x) = \begin{cases} 0 & \text{for } x < a_1 \\ r_{l_A}(x) & \text{for } a_1 \leq x < a_2 \\ 1 & \text{for } a_2 \leq x \leq a_3 \\ r_{u_A}(x) & \text{for } a_3 < x \leq a_4 \\ 0 & \text{for } x > a_4, \end{cases} \quad (1)$$

where $a_1, a_2, a_3, a_4 \in \mathbb{R}$, $a_1 \leq a_2 \leq a_3 \leq a_4$, $r_{l_A} : [a_1, a_2] \rightarrow [0, 1]$ is a nondecreasing upper semicontinuous function, and $r_{u_A} : [a_3, a_4] \rightarrow [0, 1]$ is a nonincreasing upper semicontinuous function. Functions r_{l_A} and r_{u_A} are called sometimes the left and the right “arms” (or “sides”) of the fuzzy number, respectively.

A useful notion for dealing with a fuzzy number is the so-called α -level set, also known as the α -cut. For $\alpha \in (0, 1]$ the α -level set of the fuzzy number A is the ordinary (non-fuzzy) set defined as

$$A_\alpha = \{x \in \mathbb{R} : \mu_A(x) \geq \alpha\} \quad (2)$$

and the 0-level set is usually intended to be given by $A_0 = \text{cl}(\text{supp} A)$. The family $\{A_\alpha : \alpha \in [0, 1]\}$ is a set representation of the fuzzy number A . According to the resolution identity proposed by Zadeh, we can represent the membership function as:

$$\mu_A(x) = \sup\{\alpha \cdot \mathbf{1}_{A_\alpha}(x) : \alpha \in (0, 1]\}, \quad (3)$$

where $\mathbf{1}_{A_\alpha}(x)$ denotes the indicator function of A_α .

On the basis of the notion of α -level, a fuzzy number A can be viewed as a fuzzy subset of $\mathbb{R}$ such that its α -level sets are nonempty compact and convex sets of $\mathbb{R}$, that is, nonempty compact intervals. Hence, for each $\alpha \in [0, 1]$ we have that $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, where

$$\begin{aligned} A_L(\alpha) &= \inf\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}, \\ A_U(\alpha) &= \sup\{x \in \mathbb{R} : \mu_A(x) \geq \alpha\}. \end{aligned} \quad (4)$$

If the sides of the fuzzy number A are strictly monotonic functions, then from Eq. (1) one can see easily that $A_L(\alpha)$ and $A_U(\alpha)$ are inverse functions of r_{l_A} and r_{u_A} , respectively.

In statistical analysis of random data we use functions (and, in particular, operations) of the observed random samples. These functions, called statistics, can be also defined for fuzzy random data. Their membership functions can be derived by using Zadeh's *extension principle*. This principle has the following formulation (Zadeh 1975):

Let $\mathbb{X}$ be a Cartesian product of universes $\mathbb{X} = \mathbb{X}_1 \times \dots \times \mathbb{X}_r$, and $A_1, \dots, A_r$ be r fuzzy subsets of $\mathbb{X}_1, \dots, \mathbb{X}_r$, respectively. Let f be a mapping from $\mathbb{X} = \mathbb{X}_1 \times \dots \times \mathbb{X}_r$ to a universe $\mathbb{Y}$ such that $y = f(x_1, \dots, x_r)$. The extension principle allows us to induce from the r fuzzy sets A_i a fuzzy set B on $\mathbb{Y}$ through f , $B = f(A_1, \dots, A_r)$ such that

$$\mu_B(y) = \begin{cases} \sup_{x_1, \dots, x_r | y=f(x_1, \dots, x_r)} \min\{\mu_{A_1}(x_1), \dots, \mu_{A_r}(x_r)\} & \text{if } f^{-1}(y) \neq \emptyset \\ 0 & \text{if } f^{-1}(y) = \emptyset \end{cases} \quad (5)$$

In case of $A_i, i = 1, \dots, n$ being fuzzy numbers, and f being either an injective or a continuous function, then for each $\alpha \in [0, 1]$ the α -level of $B = f(A_1, \dots, A_r)$ can be shown to be equal to $B_\alpha = (f(A_1, \dots, A_r))_\alpha$ with

$$\begin{aligned} (f(A_1, \dots, A_r))_L(\alpha) &= \min_{(x_1, \dots, x_r) \in (A_1)_\alpha \times \dots \times (A_r)_\alpha} f(x_1, \dots, x_r), \\ (f(A_1, \dots, A_r))_U(\alpha) &= \max_{(x_1, \dots, x_r) \in (A_1)_\alpha \times \dots \times (A_r)_\alpha} f(x_1, \dots, x_r). \end{aligned} \quad (6)$$

Thus, the application of the extension principle for the calculation of the membership function of $y = f(x_1, \dots, x_r)$ is equivalent to the application of the interval arithmetics on α -level sets of the arguments of this function (see Nguyen 1978 as a basis

for the proof). For instance, if A and B are fuzzy numbers, then,

$$\begin{aligned} (A + B)_L(\alpha) &= A_L(\alpha) + B_L(\alpha), \\ (A + B)_U(\alpha) &= A_U(\alpha) + B_U(\alpha), \\ (\lambda \cdot A)_L(\alpha) &= \begin{cases} \lambda \cdot A_L(\alpha) & \text{if } \lambda \geq 0 \\ \lambda \cdot A_U(\alpha) & \text{if } \lambda < 0 \end{cases} \\ (\lambda \cdot A)_U(\alpha) &= \begin{cases} \lambda \cdot A_U(\alpha) & \text{if } \lambda \geq 0 \\ \lambda \cdot A_L(\alpha) & \text{if } \lambda < 0 \end{cases} \end{aligned}$$

for any $\lambda \in \mathbb{R}$, and $\alpha \in [0, 1]$.

Fuzzy Random Variables

Uncertainty, understood as randomness, is well described in Probability Theory. The concept of a random variable, which is basic in this theory, is well-known and its definition does not need to be recalled in this article.

However, when we observe random experimental data which are imprecise, a useful tool to model either the imprecise perception of values coming from real-valued random variables or the random mechanisms generating directly these imprecise data is the one associated with the so-called concept of fuzzy random variables. Actually, we can consider two different approaches to the concept of fuzzy random variable; the motivation for these approaches and the situations they apply to are different, but the formalization of the second notion and the associated statistical methodology can be applied to the first one.

Historically, the first widely accepted definition of the fuzzy random variable was proposed by Kwakernaak (1978, 1979). Kruse (1982) proposed an interpretation of this notion, and according to this interpretation a fuzzy random variable Z may be considered as a fuzzy perception of an unknown true real-valued random variable Z_0 associated with a random experiment, and referred to as 'the original' of Z . Below, we recall the version of this definition elaborated by Kruse and Meyer (1987).

Definition 2 (Kruse and Meyer 1987) Let $(\Omega, \mathcal{A}, P)$ be a probability space, where Ω is the set of all possible outcomes of a random experiment, $\mathcal{A}$ is a σ -field of subsets of Ω (the set of all possible events of interest), and P is a probability measure associated with $(\Omega, \mathcal{A})$.

A mapping $X : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$, where $\mathcal{F}_c(\mathbb{R})$ is the space of all fuzzy numbers, is called a **fuzzy random variable** if it satisfies the following properties:

- (i) $\{X_\alpha(\omega) : \alpha \in [0, 1]\}$, where $X_\alpha(\omega) = (X(\omega))_\alpha$ is a set representation of $X(\omega)$ for all $\omega \in \Omega$;

for each $\alpha \in [0, 1]$ both $X_\alpha^L : \Omega \rightarrow \mathbb{R}$ and $X_\alpha^U : \Omega \rightarrow \mathbb{R}$, with $X_\alpha^L(\omega) = \inf X_\alpha(\omega)$ and $X_\alpha^U(\omega) = \sup X_\alpha(\omega)$, are usual real-valued random variables associated with $(\Omega, \mathcal{A}, P)$.

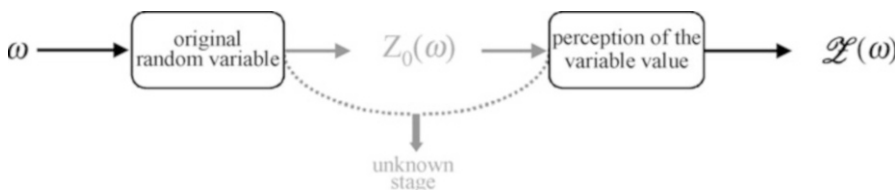
Values of a fuzzy random variable in Definition 2 have been conceived to model fuzzy perceptions of existing real-valued values (the values of the original, see Fig. 1). For instance, when we qualify the price of a given item in a specific store, we can perceive/label it as being ‘cheap’, but there is an existing price (although assumed to be unknown for the person receiving and processing data information).

Puri and Ralescu (Puri and Ralescu 1986) introduced the concept of the also called fuzzy random variable as a generalization of the concept of random set or set-valued random element (and hence, as a generalization also of the concept of random variable). According to this definition a fuzzy random variable Z may be considered as a random element associating with each experimental outcome a value which is intrinsically fuzzy. Below, we recall Puri and Ralescu’s definition.

Definition 3 (Puri and Ralescu 1986) Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$ is said to be a *fuzzy random variable* (also referred to as *random fuzzy set*) if for each $\alpha \in [0, 1]$ the set-valued mapping $X_\alpha : \Omega \rightarrow \mathcal{K}_c(\mathbb{R})$, where $\mathcal{K}_c(\mathbb{R})$ is the class of the nonempty compact intervals and $X_\alpha(\omega) = (X(\omega))_\alpha$ for all $\omega \in \Omega$, is a compact convex random sets (that is, a Borel-measurable mapping with respect to the Borel σ -field generated by the topology associated with the Hausdorff metric on $\mathcal{K}_c(\mathbb{R})$).

Remark 1 The notion of fuzzy random variable has been in fact introduced in a more general way by considering as the codomain the space of fuzzy sets of the p -dimensional Euclidean space, or even more general Banach spaces, whose α -levels are nonempty compact subsets of this space. In case one constrains to $p = 1$ and fuzzy sets being convex, then one gets the last definition.

Remark 2 Although motivation to introduce fuzzy random variables was different in the approaches by Kwakernaak/Kruse and Meyer and by Puri and Ralescu, one can prove that the notion in Definition 3 implies the one in Definition 2. As a consequence, probabilistic ideas and results for the notion in Definition 3 (or for the more general one in Remark 1) apply to the notion in Definition 2, and the same happens for statistical developments. However, many probabilistic conclusions, and most of the statistical procedures for Definition 2 are based on the assumption of having an unknown but existing original, and considering Zadeh’s extension principle, so that these conclusions and procedures are not usually applicable to deal with data coming from fuzzy random variables in Definition 3.



Statistics with Imprecise Data, Fig. 1 Fuzzy random variables in Kwakernaak/Kruse and Meyer’s sense

Remark 3 The concept of fuzzy random variable in Definition 3 can be alternatively formalized as a Borel-measurable mapping with respect to the Borel σ -field generated by the topology associated with some metrics on the space $\mathcal{F}_c(\mathbb{R})$, among them an operational one we will later refer to. Borel-measurability allows us to guarantee that notions like those of the induced distribution by a fuzzy random variable, independence of fuzzy random variables, identically distributed fuzzy random variables, and so on, can be immediately formalized in the probabilistic setting.

Values of a fuzzy random variable in Definition 3 have been conceived to model existing fuzzy values (see Fig. 2). For instance, when we classify a client of a bank in accordance with the degree of aversion to investment as having a ‘rather high’ degree, there is no underlying real-valued degree, but the classification itself is essentially imprecise.

Statistical Analysis of Fuzzy Data Corresponding to Fuzzy Perceptions of Existing Real-Valued Data

Fuzzy Estimation

When imprecise statistical data correspond to fuzzy perceptions of unobserved/unknown precise (i. e. crisp) statistical data we can treat them as observed values of fuzzy random variables in the sense of Kwakernaak/Kruse and Meyer. In such a case we can analyze imprecise data in terms of probability distributions of their unobserved originals in a similar way as precise statistical data are analyzed using methods of traditional mathematical statistics. The only difference stems from the fact that having imprecise input information in the form of fuzzy data we cannot precisely evaluate the characteristics of the

underlying probability distribution. Therefore, instead of finding precise values of the estimators of the ‘parameters’ describing the underlying probability distribution (the one of the original), it seems more coherent finding their imprecise fuzzy perceptions.

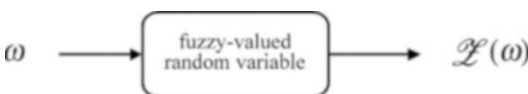
Assume that we observe a fuzzy random sample $X_1, \dots, X_n$ which is viewed as a fuzzy perception of an unobserved random sample $X_1, \dots, X_n$. Let $F(x; \theta)$ be the cumulative probability function of the original random variable X characterized by a crisp parameter $\theta \in \Theta$. Suppose now that an estimator of θ which is given by a statistic $\hat{\theta} = \phi(X_1, \dots, X_n)$ is considered. By using Zadeh’s extension principle we can consider as a *fuzzy estimator* of θ based on $X_1, \dots, X_n$ the one associating with each fuzzy sample information $(\tilde{x}_1, \dots, \tilde{x}_n)$ the *fuzzy estimate* $\phi(\tilde{x}_1, \dots, \tilde{x}_n)$ given by

$$\mu_{\phi(\tilde{x}_1, \dots, \tilde{x}_n)}(t) = \begin{cases} \sup_{(x_1, \dots, x_n) | t = \phi(x_1, \dots, x_n)} \min \{ \mu_{\tilde{x}_1}^-(x_1), \dots, \mu_{\tilde{x}_n}^-(x_n) \} & \\ \text{if } t \in \text{Im}(\phi(X_1, \dots, X_n)) & 0 \text{ otherwise} \end{cases} \quad (7)$$

Alternatively, $\phi(\tilde{x}_1, \dots, \tilde{x}_n)$ can be viewed as a fuzzy estimate of the induced fuzzy parameter $\vartheta = \theta(X)$ with $\theta(X)(t) = \sup_{X \in \mathcal{E}(\Omega, \mathcal{A}, P) | \theta(X)=t} \inf_{\omega \in \Omega} \mu_{X(\omega)}(t)$ and $\mathcal{E}(\Omega, \mathcal{A}, P)$ being the class of all possible originals of X . In many practical cases, when $\phi(x_1, \dots, x_n)$ has a simple form, the calculation of (7) is straightforward. For example, when $\phi(x_1, \dots, x_n) = \text{sample mean} = (x_1 + \dots + x_n)/n$ the α -levels of the estimate of its expected value θ are expressed as

$$(\phi(\tilde{x}_1, \dots, \tilde{x}_n)) = \left[\frac{1}{n} \sum_{i=1}^n (\tilde{x}_i)_L(\alpha), \frac{1}{n} \sum_{i=1}^n (\tilde{x}_i)_U(\alpha) \right] \quad (8)$$

Analogously, whenever the estimator of θ is given by a continuous function or an injective function, the α -levels becomes also rather simple. In other cases, the limits of the α -levels should



Statistics with Imprecise Data, Fig. 2 Fuzzy random variables in Puri and Ralescu’s sense

often be found by solving nonlinear mathematical programming problems defined by (7).

A similar approach may be applied when we try to construct fuzzy versions of the confidence intervals of the unknown parameter θ . Let us assume that we are able to find confidence intervals of θ using precise (crisp) statistical data. For example, let $[\underline{\pi}_l, +\infty)$, where $\underline{\pi}_l = \underline{\pi}_l(X_1, \dots, X_n; \delta)$, be the one-sided confidence interval for the parameter θ on the confidence level $1 - \delta$. Kruse and Meyer (1987) have shown that when we replace $\underline{\pi}_l$ with the lower limits of the α -cuts of its fuzzy version we obtain a proper confidence interval for the fuzzy perception of θ . These lower limits can be found for different values of $\alpha \in (0, 1]$ from the formula [15]

$$\begin{aligned} \underline{\Pi}_\alpha^L &= \underline{\Pi}_\alpha^L(X_1, \dots, X_n; \delta) \\ &= \inf \left\{ t \in \mathbb{R} : \forall i \in \{1, \dots, n\} \exists x_i \in X_{i,\alpha} \right. \\ &\quad \left. \text{such that } \underline{\pi}_l(x_1, \dots, x_n; \delta) \leq t \right\} \end{aligned} \quad (9)$$

where $X_{i,\alpha}$, $i = 1, \dots, n$ are respective α -cuts of the observed fuzzy data.

In a similar way we can define a fuzzy equivalent of the one-sided confidence interval $(-\infty, \bar{\pi}_u]$:

$$\begin{aligned} \bar{\Pi}_\alpha^U &= \bar{\Pi}_\alpha^U(X_1, \dots, X_n; \delta) \\ &= \sup \left\{ t \in \mathbb{R} : \forall i \in \{1 \dots n\} \exists z_i \in X_{i,\alpha} \right. \\ &\quad \left. \text{such that } \bar{\pi}_u(z_1, \dots, z_n; \delta, \geq t) \right\} \end{aligned} \quad (10)$$

where $\bar{\pi}_u(z_1, \dots, z_n; \delta) = \underline{\pi}_l(z_1, \dots, z_n; 1 - \delta)$. Moreover, exactly the same approach can be applied when we look for two-sided confidence intervals.

Summing up this section we can say that when imprecise observations may be treated as fuzzy perceptions of precise but unobserved realizations

of ordinary random variables the problem of point and interval estimation of the unknown parameters of the underlying probability distribution can be reduced to finding fuzzy versions of the formulae known from traditional mathematical statistics.

Fuzzy Statistical Tests

Testing statistical hypotheses is the second main branch of mathematical statistics. Tests of statistical hypotheses have to be applied if we want to make decisions based on the analysis of random data. When our decisions depend on the values of the parameters of probability distributions that describe observed statistical data we use parametric statistical methods. In such a case we test statistical hypotheses about the values of the parameters of probability distributions utilizing a well known equivalence between the set of values of the considered probability distribution parameter for which the null hypothesis is accepted and a certain confidence interval for this parameter. Kruse and Meyer (1987) have shown that the same equivalence exists in the case of statistical tests with fuzzy data.

Let $X_1, \dots, X_n$ denote a fuzzy sample, i. e. a fuzzy perception of the usual random sample $X_1, \dots, X_n$, from the population with the distribution P_θ . Let δ be a given number from the interval $(0, 1)$. Grzegorzewski (2000) proposed the following definition of the *fuzzy test* for vague data:

Definition 4 A function $\varphi : (\mathcal{F}_c(\mathbb{R}))^n \rightarrow \mathcal{F}(\{0, 1\})$ is called a fuzzy test for the hypothesis H , at the significance level δ , if

$$\sup_{\alpha \in [0, 1]} P\{\omega \in \Omega : \varphi_\alpha(X_1(\omega), \dots, X_n(\omega)) \subseteq \{1\} | H\} \leq \delta, \quad (11)$$

where φ_α is the α -level set (α -cut) of $\varphi(X_1, \dots, X_n)$.

The fuzzy test defined above can be regarded as a family of classical tests $\{\varphi_\alpha : \alpha \in (0, 1]\}$ for which the significance level is given as the upper bound of type I error for the whole family $\{\varphi_\alpha : \alpha \in (0, 1]\}$.

In order to give an example of a fuzzy statistical test let us consider a following simple null hypothesis: $H : \theta = \theta_0$, against the composite two-sided alternative: $K : \theta \neq \theta_0$. Suppose we know a two-sided symmetrical confidence interval $[\pi_1, \pi_2]$ for θ , on a confidence level $1 - \delta$, where $\pi_1 = \pi_1(X_1, \dots, X_n; \delta/2)$ and $\pi_2 = \pi_2(X_1, \dots, X_n; \delta/2)$ are the limits of the ordinary two-sided confidence interval. The fuzzy equivalent of this confidence interval can be calculated using the α -cuts $\Pi_\alpha = [\Pi_\alpha^L, \Pi_\alpha^U]$ for all $\alpha \in (0, 1]$, where the limits of these α -cuts can be computed from Eqs. (9) and (10) by replacing δ with $\delta/2$. The fuzzy two-sided statistical test for $H : \theta = \theta_0$, against $K : \theta \neq \theta_0$, on the significance level δ , has been defined by Grzegorzewski (2000) as a function $\varphi : (\mathcal{F}_c(\mathbb{R}))^n \rightarrow \mathcal{F}(\{0, 1\})$ with following α -cuts

$$\varphi_\alpha(X_1, \dots, X_n) = \begin{cases} \{0\} & \text{if } \theta_0 \in (\Pi_\alpha \setminus (-\Pi)_\alpha), \\ \{1\} & \text{if } \theta_0 \in ((-\Pi)_\alpha \setminus \Pi_\alpha), \\ \{0, 1\} & \text{if } \theta_0 \in (\Pi_\alpha \cap (-\Pi)_\alpha), \\ \emptyset & \text{if } \theta_0 \notin (\Pi_\alpha \cup (-\Pi)_\alpha), \end{cases} \quad (12)$$

Similarly we may obtain fuzzy tests for one-sided hypotheses using the one-to-one correspondence between the acceptance regions of the tests designated for testing these hypotheses on the significance level δ and one-sided confidence intervals for the parameter θ on the confidence level $1 - \delta$.

Grzegorzewski (2000) has shown that the membership function of the fuzzy test for the hypothesis H against K is given by

$$\begin{aligned} \mu_\varphi(t) &= \mu_\Pi(\theta_0)I_{\{0\}}(t) + \mu_{-\Pi}(\theta_0)I_{\{1\}}(t) \\ &= \mu_\Pi(\theta_0)I_{\{0\}}(t) + (1 - \mu_\Pi(\theta_0))I_{\{1\}}(t), \\ &\quad t \in \{0, 1\}, \end{aligned} \quad (13)$$

where Π is a fuzzy acceptance region depending on the considered hypotheses. Thus, the fuzzy fuzzy test defined by Eq. (12), contrary to the classical crisp test, does not lead to the binary decision – to accept or to reject the null hypothesis – but to a fuzzy decision. One may get $\mu_\varphi(0) =$

$1, \mu_\varphi(1) = 0$ which indicates that we should accept H , or $\mu_\varphi(0) = 0, \mu_\varphi(1) = 1$ which means the rejection of H . However, one may also get $\mu_\varphi(0) = \mu_0, \mu_\varphi(1) = 1 - \mu_0$, where $\mu_0 \in (0, 1)$, which can be interpreted as a degree of conviction that we should accept (μ_0) or reject ($1 - \mu_0$) the hypothesis H . Thus, in situation when μ_0 is neither 0 nor 1, a user must decide using other criteria whether to reject or to accept the considered hypothesis. There exist several approaches that are suitable for solving this problem. One of these approaches which is formulated in the language of the possibility theory has been proposed by Hryniewicz (2006) who used the results of Dubois et al. (2002) who proposed to use statistical confidence intervals of parameters of probability distributions for the construction of possibility distributions of these parameters. According to their approach, the family of two-sided confidence intervals

$$[\pi_L(x_1, \dots, x_n; 1 - \delta/2), \pi_U(x_1, \dots, x_n; 1 - \delta/2)], \quad \delta \in (0, 1) \quad (14)$$

forms the *possibility distribution* $\tilde{\vartheta}$ of the estimated value of the unknown parameter ϑ . In a similar way it is possible to construct one-sided possibility distributions based on one-sided nested confidence intervals. Hryniewicz (2006) proposed to compare this possibility distribution with a hypothetical value of the tested parameter. For this purpose he proposed to use the necessity of strict dominance measure introduced by Dubois and Prade (1983) for measuring the necessity of the strict dominance relation $\tilde{A} \succ \tilde{B}$, where $\tilde{A}$ and $\tilde{B}$ are fuzzy sets. This measure, called the *Necessity of Strict Dominance index (NSD)*, is defined as

$$\begin{aligned} NSD &= \text{Ness}(\tilde{A} \succ \tilde{B}) \\ &= 1 - \sup_{x, y; x \leq y} \min\{\mu_A(x), \mu_B(y)\}. \end{aligned} \quad (15)$$

Hryniewicz (2006) has shown that in the classical case of precise statistical data and precisely defined statistical hypotheses the value of the *NSD* index is equal to the *p*-value of the test.

In case of fuzzy data the confidence intervals used for the construction of the possibility distribution of the estimated parameter θ can be replaced by their fuzzy equivalents presented in the previous sections of this article. In his paper Hryniewicz (2006) assumes that the value of the significance level of the corresponding statistical test δ is equal to the possibility degree α that defines the respective α -cut of the possibility distribution of $\tilde{\theta}$. He also assumes that in the possibilistic analysis of statistical tests on the significance level δ we should take into account only those possible values of the fuzzy sample whose possibility is not smaller than δ . Thus, the α -cuts of the membership function $\mu_F(\theta)$ denoted by $[\mu_{F,L}^{(\alpha)}, \mu_{F,U}^{(\alpha)}]$ are equivalent to the α -cuts of the respective fuzzy confidence intervals on a confidence level $1 - \alpha$. Having the possibility distribution of the test statistic we can use Eq. (15) for the calculation of the degree on necessity that the considered statistical hypothesis has to be accepted. When we set a critical value for this characteristic we arrive at unequivocal (crisp) decisions. It is worthy to note that this approach has been generalized in (Hryniewicz 2006) to the case of testing imprecisely defined hypothesis using fuzzy statistical data.

In the previous paragraphs we have presented fuzzy statistical tests when the class the underlying probability distribution belongs to is known. Verification of this assumption when the available statistical data are imprecise may be very difficult indeed. Therefore, it would be advisable to use fuzzy equivalents of non-parametric (distribution-free) statistical methods. Unfortunately, there exist only few papers devoted to such fuzzy tests. The most interesting result has been obtained by Denœux et al. (2005) who proposed a general methodology for the construction of fuzzy tests based on rank statistics.

Statistical analysis of fuzzy random data can be also done in the Bayesian framework. First results presenting the Bayesian decision analysis for imprecise data were given in papers by Casals et al. (1986) and Gil (1988). Other approaches have been proposed by such authors as Viertl (1987), Frühwirth-Schnatter (1993), and Taheri

and Behboodian (2001). Comprehensive Bayesian model comprising fuzzy data, fuzzy hypotheses, and fuzzy utility function has been proposed in the paper by Hryniewicz (2002).

Statistical Analysis of Existing Fuzzy-Valued Data

When imprecise statistical data correspond to intrinsic fuzzy-valued data we can treat them as observed values of fuzzy random variables in the sense of Puri and Ralescu. In the literature these data are often treated as categorical/ordinal/interval-valued ones. It should be emphasized that the model given by fuzzy random variables allows us to describe and handle these data in a more expressive scale and way (in contrast to just ranking or stating simply the interval support of the values). Thus, many statistical developments for real-valued data are based on distances/deviations between values rather than on the diversity of these values. The use of the fuzzy scale allows to consider metrics with a meaning similar to that for the real-valued case (i. e., distinguishing not only the ranks of variable values w.r.t. a certain criterion, but a physical distance between them).

The distance we will consider here is the one stated by Bertoluzza et al. (1995), so that if $A, B \in \mathcal{F}_c(\mathbb{R})$

$$D_W^\varphi(A, B) = \sqrt{\int_{[0,1]} \left[\int_{[0,1]} [f_A(\alpha, \lambda) - f_B(\alpha\lambda)]^2 dW(\lambda) \right] d\varphi(\alpha)} \quad (16)$$

with $f_A(\alpha, \lambda) = \lambda \cdot A_U(\alpha) + (1 - \lambda) \cdot A_L(\alpha)$, where

- W and φ are normalized weighted measures on $[0, 1]$ formalized as probability measures on $([0, 1], \mathcal{B}_{[0,1]})$,
- W is associated with a non-degenerate distribution,
- φ is associated with a strictly increasing distribution function on $[0, 1]$.

Remark It should be remarked on D_W^φ that

- W and φ have no stochastic meaning.
- To consider W is equivalent to consider a measure weighting points 0, 1 and a certain $t_0(W) \in (0, 1)$; in case W is symmetric, then $t_0(W) = .5$.
- For each α , the choice of W allows us to weight
 - the effect of the distance between the “widths” of the α -levels (i. e., effect of the “shape” difference),
 - in comparison with the effect of the distance between their $t_0(W)$ -points (i. e., effect of the “location” difference).
- The choice of φ allows us to weight the influence of each level (i. e., degree of “imprecision”, “consensus”, “subjectivity”,...).
- D_W^φ is a versatile and operational in statistical developments with fuzzy numbers, and it behaves especially well when we consider least-squares approaches.

Since the concept of fuzzy random variable in Puri and Ralescu’s sense has been properly stated in a probabilistic context as a random element (i. e., as a Borel-measurable function), concepts like independent and identically distributed fuzzy random variables make immediate sense. Furthermore, all the main ideas, aims and concepts, and several developments can be immediately considered to deal with fuzzy data when coming from fuzzy-valued Borel-measurable mappings. In this respect, notions like either unbiasedness or consistency of a “point” (fuzzy- or real- valued) estimator of a (fuzzy- or real- valued) ‘parameter’ associated with the distribution of the fuzzy random variable, or the p -value and power of a test concerning such a ‘parameter’, make the same sense as in the classical case.

Several developments have been made in connection with both estimation and testing of fuzzy- and real-valued parameters associated with the distribution of a fuzzy random variable in Puri and Ralescu’s sense. We are now just recalling a few results concerning the “point” fuzzy estimation and testing of the population mean of a fuzzy random variable, which is formalized as follows:

Definition 5 (Puri and Ralescu 1986) Given a probability space $(\Omega, \mathcal{A}, P)$ and an associated fuzzy random variable $X : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$ such that $\max\{|\inf X_0|, |\sup X_0|\}$ is integrable, then, the **fuzzy expected value** (or **fuzzy mean**) of X is the fuzzy number $\tilde{E}(X|P) \in \mathcal{F}_c(\mathbb{R})$ such that for all $\alpha \in [0, 1]$

$$\begin{aligned} (\tilde{E}(X|P))_\alpha &= \text{Aumann integral of } X_\alpha \\ &= \left\{ E(X|P) | X : \Omega \rightarrow \mathbb{R}, \right. \\ &\quad \left. X \in L^1(\Omega, \mathcal{A}, P), X \in X_\alpha \text{ a.s.} \right. \\ &\quad \left. [P] \right\} \\ &= \left[E(\inf X_\alpha | P), \right. \\ &\quad \left. E(\sup X_\alpha | P) \right]. \end{aligned}$$

Remark 5 The fuzzy mean satisfies that

- If $X(\Omega) = \{\tilde{x}_1, \dots, \tilde{x}_m, \dots\} \subset \mathcal{F}_c(\mathbb{R})$, then, it is coherent with fuzzy arithmetic, that is,

$$\begin{aligned} \tilde{E}(X|P) &= P(\{\omega \in \Omega | X(\omega) = \tilde{x}_1\}) \cdot \tilde{x}_1 + \dots \\ &\quad + P(\{\omega \in \Omega | X(\omega) = \tilde{x}_m\}) \cdot \tilde{x}_m + \dots \end{aligned}$$

- Strong Laws of Large Numbers are satisfied for different metrics (like D_W^φ , and stronger ones), which also corroborates the suitability of the defined fuzzy mean as the stochastic limit of the sample ones.
- $\tilde{E}(X|P)$ is the “Fréchet expectation” of X w.r.t. D_W^φ , i. e., for all $A \in \mathcal{F}_c(\mathbb{R})$:

$$E\left(\left[D_W^\varphi(X, \tilde{E}(X|P))\right]^2 | P\right) \leq E\left(\left[D_W^\varphi(X, A)\right]^2 | P\right).$$

The interpretation of the fuzzy mean becomes more clear if we notice that for interval-valued random variables (i. e., random intervals) the expected value is given in a form of an interval with the lower limit equal to the expected value of the lower limits of observed random variables, and the upper limit equal to the expected value

of the upper limits of observed random variables. Therefore, the expected value of the fuzzy random variable can be given as a nested set of such intervals represented by respective α -cuts, as it is formally described in the definition.

The definition of other characteristics describing fuzzy random variables may be much more complicated. For example, the definition of the variance requires the introduction of the “Fréchet expectation” (defined above), as it was proposed by Körner and Näther (2002).

Fuzzy Estimation

Assume that we observe a fuzzy simple random sample $X_1, \dots, X_n$ which is viewed now as an n -tuple of n independent fuzzy random variables which are identically distributed as X . The associated **fuzzy sample mean** is the statistic given by

$$\bar{X}_n = \frac{1}{n} \cdot [X_1 + \dots + X_n].$$

Then, one can prove that (see Lubiano 1999)

Theorem 1 *The fuzzy sample mean satisfies that*

- (i) $\bar{X}_n[\cdot]$ in an “unbiased fuzzy-valued estimator” of $\tilde{E}(X|P)$ (in the sense of the fuzzy expected value defined by Puri & Ralescu). For most of the metrics we can consider, it is also a ‘strongly consistent’ fuzzy-valued estimator of $\tilde{E}(X|P)$.
- (ii) One can quantify the mean squared-type error in the fuzzy estimation by considering the real-valued expected value $E\left(\left[D_W^\varphi(\bar{X}_n, \tilde{E}(X|P))\right]^2\right)$.

Actually, Lubiano et al. (1999, 2000) proposed several developments in connection with the estimation and testing about the D_W^φ -mean squared error associated with the estimation of the population fuzzy means by means of the sample one.

Statistical Tests

One of the problems which has received a deep attention in connection with testing from fuzzy random variables in Puri and Ralescu’s sense is

that concerning two-sided tests about the mean of a fuzzy random variable (one-sample case). In order to define this problem one can decide which metrics could be used for measuring the distance between the hypothetical and observed fuzzy values of considered characteristics of the fuzzy random variables. In the case of the metrics defined by Eq. (16), the problem can be formalized as follows:

Given n independent observations from X , $X_1, \dots, X_n$, we wish to test the null hypothesis $H_0 : \tilde{E}(X|P) = A \in \mathcal{F}_c(\mathbb{R})$, which can be equivalently expressed as $H_0 : D_W^\varphi(\tilde{E}(X|P), A) = 0$.

Despite simple formulation of the testing problem the construction of statistical tests for the verification of hypotheses related to fuzzy mean values is not that simple. For example, an exact test has been developed in Montenegro et al. (2004) for so called “normal” fuzzy random variables (in Puri and Ralescu’s sense (1985)). Although the method is exact and easy-to-apply, the assumption of X being fuzzy “normal” ($X = V + \mathcal{N}(0, 1)$, with $V \in \mathcal{F}_c(\mathbb{R})$) is quite restrictive and often unrealistic, as it means that all observed fuzzy values are described by the same membership function shape with maybe different location.

In a more general case asymptotic tests based on Large Sample Theory have been developed for the same purpose. However, they are usually hardly applicable (except for some simple special cases) in practice. For example, the asymptotic distribution of the statistic proposed by Körner (2000) involves unknown parameters such as correlations between random variables describing membership functions of observed fuzzy variables. In the paper by Montenegro et al. (2004) devoted to the problem of testing the equality of two fuzzy mean values, it is assumed that X takes on a finite number of values, and large sample sizes would be required anyway in order to estimate unknown parameters of the model.

Simulation studies have been considered to analyze the extent and applicability of the asymptotic test by Montenegro et al. (2004). These studies have confirmed that in estimating the unknown parameters (the eigenvalues of a certain

correlation matrix) entails a substantial loss of precision w.r.t. the nominal significance level. Based on this empirical conclusion, the use of D_W^φ and the Generalized Bootstrapped CLT (Giné and Zinn 1990) allow us to consider bootstrap techniques in this context.

The bootstrap technique consists in taking random samples (i. e. re-sampling) from the original one and calculating the value of the considered statistic. By repeating this procedure many times we obtain the sample distribution of the test statistic which may be thus used for testing purposes. In case of fuzzy random data we get the following method proposed by González-Rodríguez et al. (2006):

Theorem 2 *Given a fuzzy random variable $X : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$ associated with the probability space (Ω, A, P) and such that*

- $\max\left\{(\inf X_0)^2, (\sup X_0)^2\right\}$ is integrable
- $X_1, \dots, X_n$ are i.i.d. as X
- $X_1^*, \dots, X_n^*$ is a bootstrap sample from $X_1, \dots, X_n$

then, to test H_0 at the nominal significance level $\alpha \in [0, 1]$, H_0 should be rejected whenever

$$\frac{[D_W^\varphi(\overline{X}_n, U)]^2}{\hat{S}_n^2} > z_\alpha,$$

where z_α is the $100(1 - \alpha)$ fractile of the bootstrap distribution of

$$T_n = [D_W^\varphi(\overline{X}_n^*, \overline{X}_n)]^2 / \hat{S}_n^{*2}$$

with

$$\begin{aligned} \overline{X}_n^* &= \sum_{i=1}^n X_i^* / n, \quad \hat{S}_n^{*2} \\ &= \sum_{i=1}^n [D_W^\varphi(X_i^*, \overline{X}_n^*)]^2 / (n - 1). \end{aligned}$$

Other bootstrap tests about means of fuzzy random variables which have been recently developed are the following ones:

- One-sided hypotheses tests in the one-sample case.
- Tests for the equality of means of two FRVs
 - for two independent samples
 - for two linked samples.
- Tests for the equality of means of J FRVs (ANOVA):
 - for J independent samples.

From comparative extensive simulation studies that have been performed recently at the University of Oviedo and the European Center for Soft Computing we can draw the following practical conclusions:

- For small/medium samples, the bootstrap method performs and behaves much better than the asymptotic one.
- For large sample sizes (over 300), the improvement is not that remarkable, but the bootstrap approach still provides the best approximation to the nominal significance level.

Taking into account that fuzzy statistical test for testing hypotheses about other characteristics of fuzzy random variables are even more complicated than the procedures designed for testing the hypotheses about mean values we can claim that the bootstrap technique is probably the most promising for dealing with random fuzzy observations.

Future Directions

Statistical analysis of imprecise data is still a developing area of science. Future directions of its development are tightly connected with the development of methods that may be used for the description of uncertainties of different types. It has been pointed out by P. Walley (see, e. g. (Walley 1996) for a good overview) that traditional probability is not sufficient for good description of different types of uncertainty. Different mathematical models, such as e. g. Dempster–Shafer belief functions, possibility distributions, lower and upper probabilities,

lower and upper previsions, and many others, have been proposed for this purpose. However, for the most general models describing uncertainty appropriate statistical methods have not been proposed yet. Therefore, statistical methods for handling very general imprecise data have to be developed in the future.

Another, but much more specific, future direction for the development of statistical analysis of imprecise data is related to the analysis of intrinsically fuzzy data. In contrast to the situation when fuzzy observations may be considered as fuzzy perceptions of real-valued observations, many notions known from traditional statistics are still waiting for their widely accepted definitions, and statistical methods of analysis.

The most challenging future direction is related to Zadeh's paradigm of "computing with words". First of all, we need operational methods for the representation of linguistic concepts which could be useful is statistical analysis of imprecisely reported (with words!) statistical data. Moreover, we also need methods for convincing presentation of the results of computations to users who have only limited knowledge of mathematics and statistics.

Bibliography

Primary Literature

- Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer, Dordrecht
- Bertoluzza C, Corral N, Salas A (1995) On a new class of distances between fuzzy numbers, mathware and soft computing, vol 2. Departament of Computer Science and Artificial Intellingence of the University of Granada and Secció de Matemàtiques i Informàtica of the Universitat Politècnica de Catalunya, Granada, Barcelona, pp 71–84. <http://docto-si.ugr.es/Mathware/ENG/mathware.html>
- Casals R, Gil MA, Gil P (1986) The fuzzy decision problem. An approach to the problem of testing statistical hypotheses with fuzzy information. *Eur J Oper Res* 27: 371–382
- Deneoux T, Masson M-H, Hébert PA (2005) Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets Syst* 153:1–28
- Dubois D, Prade H (1978) Operations on fuzzy numbers. *Int J Syst Sci* 9:613–626
- Dubois D, Prade H (1980) Fuzzy sets and systems. In: *Theory and applications*. Academic, New York
- Dubois D, Prade H (1983) Ranking fuzzy numbers in the setting of possibility theory. *Inf Sci* 30:184–244
- Dubois D, Prade H (1988) Possibility theory. Plenum Press, New York
- Dubois D, Foulloy L, Mauris G, Prade H (2002) Probability-possibility transformations, triangular fuzzy-sets and probabilistic inequalities. In *Proceedings of the ninth international conference IPMU*, Annecy, pp 1077–1083
- Frühwirth-Schatter S (1993) Fuzzy Bayesian inference. *Fuzzy Sets Syst* 60:41–58
- Gil MÁ, Kacprzyk J, Fedrizzi M (1988) Probabilistic-possibilistic approach to some statistical problems with fuzzy experimental observations. In: *Combining fuzzy imprecision with probabilistic uncertainty in decision making*. Springer, Berlin, pp 286–306
- Gil MÁ, López-Díaz M, Ralescu DA (2006a) Overview on the development of fuzzy random variables. *Fuzzy Sets Syst* 157:2546–2557
- Giné E, Zinn J (1990) Bootstrapping general empirical measures. *Ann Probab* 18:851–869
- González-Rodríguez G, Montenegro M, Colubi A, Gil MA (2006) Bootstrap techniques and fuzzy random variables. Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets Syst* 157:2608–2613
- Grzegorzewski P (2000) Testing statistical hypotheses with vague data. *Fuzzy Sets Syst* 112:501–510
- Hryniewicz O (2006) Possibilistic decisions and fuzzy statistical tests. *Fuzzy Sets Syst* 157:2665–2673
- Hryniewicz O, Grzegorzewski P, Gil MÁ (2002) Possibilistic approach to the Bayes statistical decisions. In: *Soft methods in probability, statistics and data analysis*. Physica, Heidelberg, pp 207–218
- ISO/IEC (1995) Guide to the expression of uncertainty in measurement (GUM). ISO/IEC, Geneva
- Körner R (2000) An asymptotic α -test for the expectation of random fuzzy variables. *J Statist Plann Inference* 83: 331–346
- Körner R, Näther W (2002) On the variance of random fuzzy variables. In: Bertoluzza C, Gil MÁ, Ralescu DA (eds) *Statistical modeling, analysis and management of fuzzy data*. Physica, Heidelberg, pp 25–42
- Kruse R (1982) The strong law of large numbers for fuzzy random variables. *Inf Sci* 28:233–241
- Kruse R, Meyer KD (1987) Statistics with vague data. D. Riedel Publishing Company, Dordrecht
- Kwakernaak H (1978) Fuzzy random variables. Definitions and theorems. *Inf Sci* 15:1–15
- Kwakernaak H (1979) Fuzzy random variables. Algorithms and examples for the discrete case. *Inf Sci* 17: 253–278
- Lubiano MA (1999) Medidas de variación de elementos aleatorios. Ph D thesis, University of Oviedo
- Lubiano MA, Gil MA (1999) Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Stat Pap* 40:277–295
- Lubiano MA, Gil MA, López-Díaz M, López-García MT (2000) The $\bar{\lambda}$ -mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets Syst* 111:307–317

- Montenegro M, Colubi A, Casals MR, Gil MA (2004) Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* 59:31–49
- Nguyen HT (1978) A note on the extension principle for fuzzy sets. *J Math Anal Appl* 64:369–380
- Puri ML, Ralescu DA (1985) The concept of normality for fuzzy random variables. *Ann Probab* 13:1373–1379
- Puri ML, Ralescu DA (1986) Fuzzy random variables. *J Math Anal Appl* 114:409–422
- Taheri SM, Behboodian J (2001) A Bayesian approach to fuzzy hypotheses testing. *Fuzzy Sets Syst* 123:39–48
- Terán P (2007) Probabilistic foundations for measurement modelling with fuzzy random variables. *Fuzzy Sets Syst* 158:973–986
- Viertl R (1987) Is it necessary to develop a fuzzy Bayesian inference. In: *Probability and Bayesian statistics*. Plenum, New York, pp 471–475
- Viertl R (1996) *Statistical methods for non-precise data*. CRC Press, Boca-Raton
- Walley P (1996) Measures of uncertainty in expert systems. *Artif Intell* 114:1–58
- Zadeh LA (1956) Fuzzy sets. *Inf Control* 8:338–353
- Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. *Inf Sci* 8(1): 199–249; 8(2):301–353; 9(3):43–80

Books and Reviews

- Bertoluzza C, Gil MÁ, Ralescu DA (eds) (2002) *Statistical modeling, analysis and management of fuzzy data*. Physica, Heidelberg
- Dubois D, Lubiano MA, Prade H, Gil MÁ, Grzegorzewski P, Hryniewicz O (eds) (2008) *Soft methods for handling variability and imprecision*. Springer, Berlin
- Gil MÁ, López-Díaz M, Ralescu DA (2006b) Overview on the development of fuzzy random variables. *Fuzzy Sets Syst* 157:2546–2557
- Grzegorzewski P, Hryniewicz O, Gil MÁ (2002) Soft methods in probability. In: *Statistics and data analysis*. Physica, Heidelberg
- Kruse R, Gebhardt J, Gil MÁ (1999) Fuzzy statistics. In: Webster JG (ed) *Encyclopedia of electrical and electronics engineering*. Wiley, New York
- Lawry J, Miranda E, Bugarin A, Li S, Gil MÁ, Grzegorzewski P, Hryniewicz O (eds) (2006) *Soft methods for integrated uncertainty modeling*. Springer, Berlin
- López-Díaz M, Gil MÁ, Grzegorzewski P, Hryniewicz O, Lawry J (eds) (2004) *Soft methodology and random information systems*. Springer, Berlin
- Taheri SM (2003) Trends in fuzzy statistics. *Austral J Statist* 32:239–257



Some Remarks on the Foundations of Numerical Analysis

Ralph Baker Kearfott
Department of Mathematics, University of
Louisiana, Lafayette, LA, USA

Article Outline

Glossary
Definition of Numerical Analysis
Introduction
Cn Modern Numerical Analysis
Future Directions
References

Glossary

Accuracy The error of approximation. Absolute accuracy is the difference between the approximation and true value, while relative accuracy is that difference divided by the absolute value of the true value

Algorithms Sets of instructions that, when followed, transform a set of input items into a set of output items with given properties

Approximation theory Approximation of functions by functions that are simple to compute, such as polynomials or rational functions

Computational science Use of mathematical models employing algorithms produced from numerical analysis to study real-world phenomena. Prominent examples are weather prediction and astrophysics

Computer architecture The circuitry design for a digital computer. Modern designs allow concurrent, or simultaneous, computations to reduce the elapsed time to complete an overall task

Discretization A finite set of simple objects meant to approximate an ideal but unrepresentable object

Floating-point arithmetic Is arithmetic on a finite set of numbers characterized by a number base β , a mantissa consisting of a fixed number t of base β digits

Heuristic (as an adjective) Not always or certainly so, but assumed to usually be the case. (as a noun): a rule within an algorithm that makes an assumption that is not always true, but is useful to assume is often true

Numerical algorithms Algorithms using the approximate arithmetic of numerical analysis, to obtain solutions to problems treated by numerical analysis

Parallel computation Assignment of independent parts of an overall computation to separate processors, so the independent computations can be done at the same time, that is, concurrently. See also computer architecture

Round-off error The error in a computation resulting from approximating a real number by an expansion (decimal, binary, or other base) with a finite number of digits

Scientific computing The practice of numerical analysis, primarily in the modern digital computer age (from the late 1940s). Connotes applying algorithms produced from numerical analysis to the solution of mathematical problems arising from the sciences

Simulation The use of numerical algorithms to solve mathematical models of real-world phenomena (physical, biological, economical, etc.), to predict the future or determine consequences of particular actions

Definition of Numerical Analysis

A common definition of numerical analysis is the study of algorithms based on rounded or arithmetic for the problems of continuous mathematics. This study encompasses discrete approximations of the mathematical problem and approximation of the discretization within the adopted arithmetic

system, along with an analysis of both the discretization errors and the rounding errors within the arithmetic.

Closely related terms are scientific computing and computational science.

Introduction

Due to its ties to mathematical modeling of important real-world problems, numerical analysis is currently taught both as a core course and a service course, both at the undergraduate and graduate levels, within mathematics, engineering, and computer science curricula. During much of the second half of the twentieth century, until personal desktop and laptop computers became ubiquitous, algorithms based on numerical analysis were the primary use of digital computers. Numerical computations are still the primary use of current supercomputers, which are designed largely for that purpose.

Textbooks, monographs, and tutorials on numerical analysis can be lucrative if well-written and widely adopted. Thus, numerous authors have tried their hand, with some being popular and up-to-date.

Three pillars of numerical analysis are as follows:

Approximation theory: Approximation of functions by functions that are simple to compute, such as polynomials or rational functions.

Finite arithmetic systems: Construction and analysis of arithmetic systems approximating the real or complex numbers, but allowing easier manipulation and interpretation of results.

Algorithm development: Creation and analysis of sets of instructions for performing mathematical computations of interest.

Numerical analysis is commonly thought to have begun with the advent of modern computers in the 1950s. However, elements of approximation theory and even algorithms extend back to the ancient Greeks, analysis of finite arithmetic systems began in the Age of Enlightenment and the

metric system, and notable elements of current numerical algorithms date from the nineteenth and early twentieth centuries.

Early History

Perhaps the most famous very early approximation is Archimedes' techniques for approximating the circumference and area of a circle by chords and by areas of polygons, often referenced as Archimedes' approximation of π . In fact, the technique can be viewed as an algorithm for approximating π to a desired accuracy (see Casselman 2012).

Modern numerical analysis is closely tied to core concepts in differential and integral calculus. In fact, the difference quotient is a numerical approximation to the derivative, while Riemann sums are a type of discretization of the definite integral. Thus, Leibnitz' and Newton's first publications on the calculus (Leibniz 1684; Newton 1687) can be considered seminal works in numerical analysis.

In fact, Leibnitz went even further, proposing the design of mechanical digital calculators based on binary arithmetic, the predominant arithmetic in modern digital computers (see Leibniz 1703). Leibnitz' work, in turn, could not have happened without Fibonacci's introduction of Hindu–Arabic numerals into Europe in Fibonacci (1202), (also see the English translation Sigler (2002)).

The introduction of the metric system motivated the study of round-off error. Indeed, the various measurement units were now decimally related, and engineering and scientific computations such as with trigonometric and exponential functions could be done more efficiently with decimal approximations. In fact, we see modern round-off error analysis beginning to take shape in Gauss (1809). Rounded computations, round-off error analysis, and approximation of transcendental functions seem to have been a part of German and French high school curricula in the nineteenth century, as we see in Schnellinger (1879) and Burat (1885).

Various numerical algorithms in common use today arose during this period. Notable ones, for obtaining approximate solutions to initial value problems of ordinary differential equations,

include Euler's method, first explained in Euler (1768) and Runge-Kutta methods, in Runge (1895) and Kutta (1901); see also Butcher (1996).

A basic tool in approximation theory is Taylor polynomials, treated extensively in calculus courses. The roots of Taylor polynomials and Taylor series are in India in the late Middle Ages, while a seminal publication is Taylor (1715).

Modern History

Modern approximation theory uses tools of functional analysis developed in the nineteenth and early twentieth century, while roundoff error analysis matured from the 1950s to the present, with the advent of modern electronic digital computers. Algorithmic development accelerated during the same period, with increasing emphasis on taking advantage of concurrent computer architectures and on developing algorithms for specific important simulations.

An early but interesting effort in this modern era, Richardson (1922), was ahead of its time. Richardson proposed a model for weather prediction and proposed numerical methods for solving it. He carried out hand calculations that did not successfully reproduce the actual weather recorded in Europe, not because of his numerical methods, but because the physical assumptions underlying his model were lacking. Not anticipating modern electronic digital computers, he amusingly stated there would be no possibility of carrying out the computations before the weather actually occurred; today, however, accurate forecasts using Richardson's and others' ideas are given more than a week in advance. Richardson even proposed an algorithm for parallel computation. However, at the time "computer" meant a person hired to do hand computations; Richardson proposed a room full of such people with a director, similar to an orchestra conductor, coordinating the calculations between the individuals, an example of parallel computation before its time.

Beginning in the late 1940s, the development of scientific computing is too extensive to thoroughly or impartially cover here. Chapters in the book Nash (1990) provide a good review through

1987, while the chapter authors constitute a list of some of the most prominent figures.

Cn Modern Numerical Analysis

Numerous textbooks introduce approximation theory, computer arithmetic, and algorithms in scientific computing. A particular such book may or may not treat one of these three pillars well; hints about the emphasis can sometimes be inferred from the title, for example, "Numerical Methods" vs "Numerical Analysis." They range from, essentially, user guides for software packages to functional analysis-oriented treatment of approximation theory or discretization of systems of partial differential equations. The texts are often tied to implementations in a particular computer language. Caution has been advised when teaching from some elementary "methods" texts. For example, experts in software for the solution of systems of ordinary differential equations have advised students to consider using professional software for complex simulations, rather than implement the simplified concept-illustrating algorithms appearing in the text.

Modern numerical analysis is based on floating-point arithmetic. The most widespread floating-point system in use today is the IEEE 754 64 bit binary standard; see IEEE-754 (2019). Not all textbooks explicitly cover elements of floating-point arithmetic and their impact on rounding error, and lack of understanding can be a source of confusion for engineering students interpreting simulation results. Better understanding is needed on how particular programming languages or software display floating-point quantities, and on the connection of particular floating-point systems to classical forward and backward round-off error analysis. Failure to properly take account of round-off error or use of finite arithmetic has led to catastrophic results in automatically controlled systems. Prominent examples include the administration of fatal doses in radiation therapy for cancer and incorrect targeting in a missile defense system.

Specific applications do not need the full strength of IEEE 64 bit arithmetic. For example,

in machine learning, the underlying calculations are inherently approximate or heuristic, and execution time, manufacturing expense, and electrical power can possibly be saved by implementing an arithmetic with less significant digits and fewer facilities for exception handling.

Classical References

Some favorite classical texts and references on this author's bookshelf are Wilkinson (1963), Wilkinson (1965), Young (1971), de Boor (1978), Isaacson and Keller (1994), Ortega (1972), Cheney (1966), Strang and Fix (1973), Forsythe et al. (1977), Dongarra et al. (1979), Anderson et al. (1992), Schumaker (2007), Dennis and Schnabel (1996), and Golub and van Loan (1996). Additional references in our perspective appear in our text Ackleh et al. (2010).

Future Directions

Numerical analysis and scientific computing will continue to evolve along with advances in electronic computers and with ever more ambitious projects to model scientific, engineering, economic, and social phenomena, and along with ever more ambitious projects in automation. Although the foundations will remain the same, algorithm design and efficiency analyses may be guided more and more by

- Special purpose computer architectures
- Particular applications

References

- Ackleh AS, Allen EJ, Kearfott RB, Seshaiyer P (2010) Classical and modern numerical analysis. CRC Numerical Analysis and Scientific Computing, Chapman & Hall. CRC Press, 2000 N.W. Corporate Blvd., Boca Raton, FL 33431-9868, USA, theory, methods and practice
- Anderson E, Bai Z, Bischof C, Demmel J, Dongarra J, Croz JD, Greenbaum A, Hammarling S, McKenney A, Ostrouchov S, Sorensen D (1992) LAPACK's user's guide. Society for Industrial and Applied Mathematics. updated editions are available, Philadelphia, PA
- de Boor C (1978) A practical guide to splines. Springer, New York
- Burat E (1885) *Traité D'Arithmétique a L'usage des élèves des lycées et collèges*. Eugène Belin et Fils., Paris, found November 16, 2020 at <https://gallica.bnf.fr/ark:/12148/bpt6k202612w.texteImage>
- Butcher J (1996) A history of runge-kutta methods. *Appl Numer Math* 20(3):247–260. [https://doi.org/10.1016/0168-9274\(95\)00108-5](https://doi.org/10.1016/0168-9274(95)00108-5). URL <http://www.sciencedirect.com/science/article/pii/0168927495001085>
- Casselman B (2012) Archimedes on the circumference and area of a circle. Web page, URL <http://www.ams.org/publicoutreach/feature-column/fc-2012-02>. Accessed 18 Jan 2021. (The copyright on Archimedes' original treatise has probably expired, though)
- Cheney EW (1966) Introduction to approximation theory. McGraw-Hill, New York
- Dennis JE Jr, Schnabel RB (1996) Numerical methods for unconstrained optimization and nonlinear equations. In: *Classics in applied mathematics*, vol 16. SIAM, Philadelphia, PA, originally published by Prentice-Hall, 1983
- Dongarra JJ, Moler CB, Bunch JR, Stewart GW (1979) LINPACK users' guide. SIAM, Philadelphia
- Euler L (1768) *Institutionum Calculi Integralis*. Impensis Academiae Imperialis Scientiarum Petropoli, retrieved in PDF from <https://ia802205.us.archive.org/28/items/institutionesca1020326mbp/institutionesca1020326mbp.pdf> on 20 Jan 2021)
- Fibonacci L (1202) *Liber abaci* (book of calculation). Handwritten Manuscript, excerpts found at <https://www.maa.org/book/export/html/841546> on 20 Jan 2021
- Forsythe GE, Malcolm MA, Moler CB (1977) Computer methods for mathematical computations. Prentice-Hall Professional Technical Reference
- Gauss C (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Carl Friedrich Gauss Werke, Hamburgi sumptibus Frid. Perthes et I.H. Besser, URL <https://books.google.com/books?id=ORUOAAAAQAAJ>, eBook and PDF currently available free of charge from Google at [https://books.google.com/books/about/Theoria_motus_corporum_coelestium_in_sec.html?id=unhbox\voidb@x\hbox{\\$=\\$}ORUOAAAAQAAJ](https://books.google.com/books/about/Theoria_motus_corporum_coelestium_in_sec.html?id=unhbox\voidb@x\hbox{$=$}ORUOAAAAQAAJ)
- Golub GH, van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins University Press
- IEEE-754 (2019) IEEE 754–2019, standard for floating-point arithmetic. IEEE, New York. <https://doi.org/10.1109/IEEESTD.2019.876622>
- Isaacson E, Keller HB (1994) Analysis of numerical methods. Dover, New York. ebook available; originally published by John Wiley & Sons, 1966
- Kutta W (1901) Beitrag zur näherungsweise Integration totaler Differentialgleichungen. *Zeit Math Phys* 46: 435–453
- Leibniz G (1703) Explication de l'arithmétique binaire. *Mem l'Acad R Sci* 3:85–93

- Leibniz GW (1684) *Nova methodus pro maximis et minimis, itemque tangentibus, quae nec fractas, nec irrationales quantitates moratur, et singulare pro illis calculi genus*. *Acta Eruditorum*:467–473
- Nash SG (ed) (1990) *A history of scientific computing*. Association for Computing Machinery, New York, NY, USA
- Newton I (1687) *Philosophiae naturalis principia mathematica*. J Societatis Regiae ac Typis J. Streater, URL <https://books.google.com/books?id=dVKAQAAIAAJ>, in Latin. Retrieved in PDF f on 19 Jan 2021
- Ortega JM (1972) *Numerical analysis: a second course*. Academic Press, New York
- Richardson LF (1922) *Weather prediction by numerical process*. Cambridge University Press. URL <https://archive.org/details/weatherpredictio00richrich>
- Runge C (1895) Ueber die numerische Auflösung von Differentialgleichungen. *Math Ann* 46:167–178
- Schnellinger J (1879) Das gekürzte Rechnen (basically, truncated arithmetic or inexact arithmetic). *Zeitschrift für das Realschulwesen* 1879:257–284, (currently available in Google Books)
- Schumaker L (2007) *Spline functions: basic theory*, 3rd. Cambridge Mathematical Library, Cambridge University Press, DOI <https://doi.org/10.1017/CBO9780511618994>, first edition published by WileyInterscience, 1980 Sigler
- LELE (2002) *Fibonacci's Liber Abaci: A Translation into Modern English of Leonardo Pisano's Book of Calculation*. Sources and studies in the history of mathematics and physical sciences. Springer-Verlag, Berlin/Heidelberg/London, etc
- Strang G, Fix GJ (1973) *An analysis of the finite element method*. Prentice-Hall, Englewood Cliffs
- Taylor B (1715) *Methodus incrementorum directa et inversa*. Londini, retrieved in PDF from https://ia800900.us.archive.org/6/items/UFIE003454_TO0324_PNI-2529_000000/UFIE003454_TO0324_PNI-2529_000000.pdf on 20 Jan 2021
- Wilkinson J (1965) *The algebraic eigenvalue problem*. Clarendon Press, Oxford
- Wilkinson JH (1963) *Rounding errors in algebraic processes*. Prentice-Hall, Englewood Cliffs
- Young DM (1971) *Iterative solution of large linear systems*. Academic Press

Index

A

- Absolute set approximation regions, 522–523
- Abstract algebra, 529
- Abstraction, 223
- Acceptable-risk IP addresses, 628
- Access control, 616–617
 - vs. privacy, 618–619
- Access control and privacy protection (ACPP), 618–622
- Access control lists (ACLs), 633
- Accuracy, 360, 377, 378, 387, 911
- Actionable security events, 627
- AdaBoost algorithm, 856
- Adaptive-network-based inference system (ANFIS), 266
- Adaptive Neuro-Fuzzy Inference System (ANFIS), 834
- Adaptive resonance theory (ART), 728
- Advanced persistent threat (APT), 627, 635
- Adware, 244
- Aggregate expression, 189
- Aggregation operators, 711–715, 723
 - fuzzy integrals, 720–722
 - linear combination of order statistics and OWA operators, 718–719
 - Losonczi's means, 718
 - quasi-arithmetic means, 716–717
 - weighted means and quasi-weighted means, 717–718
 - WOWA operator, 719–720
- AI-engineering, 192
- Akaike estimation, 854
- Aleatory uncertainty, 369
- Alefeld, Götz, 705
- Algebra, 327
- σ -Algebra, 51
- Algebraic logic, 530
- Algebraic models, 530–532
 - Boolean algebras and classical logic, 532–535
 - definition, 529
 - non-Boolean algebras and non-classical logics, 535–539
- Algebra representation theorem
 - definition, 529
 - implication algebras, 536
 - pierce algebras, 537
 - positive implication algebra, 535
 - quasi-Boolean algebra, 539
- Algorithms, 911
- Alliteration, 883
- Allocation rule, 557
- Alpha-cuts, computation of, 705
- AlphaGO, 269
- Alternative minimization algorithm, 556
- America Online (AOL), 655
- Analysis of sensitivities, 66
- Anderson's Wiener process, 351
- k -Anonymity, 617
- Application programming interfaces (APIs), 489
- Approximate arithmetic, 693, 697–699
- Approximation
 - by graded dominance relation, 438
 - by multi-graded dominance relations, 440–441
 - space, 880, 881
 - theory, 911
- Apriori principle, 190
- Arithmetic expression, 29
- Arithmetic mean, 716, 722
- ARKM algorithm, 559, 560
- ARMAX models, 736
- Artificial intelligence (AI), 178, 181, 712, 727
- Artificial neural network, 266
- Artificial neurons, 266
- n -Ary relation, 343
- Association rules, 231, 232, 451–452
- Atom, 506
- Atomic descriptions, 260
- Atomic infinite structures, 509
- Atomicity, consistence, isolation and duration (ACID), 180
- Atomless structures, 509
- Attribute(s), 421, 523, 878, 882–885
 - domain of, 523
 - reduction and significance of, 526
 - reducts, 526–527
 - significance of, 527
- Attribute-based access control (ABAC), 616, 619, 622, 642
- Attribute-based broadcast group key management (AB-BGKM), 643
- Attribute-based data model (ABDM), 150
- Attribute-value blocks, 390
- Austin Public Library, 494

Australian Credit, 306
 Authorization Diagram (AD), 677, 681, 682, 684
 Automated underwater mapping, 705
 Automatic theorem proving, 703
 Automorphism, 507
 Average precisions, 321
 Axiomatic approach, 865
 Axiomatizable, 502
 Axiom of choice, 332
 Axioms of topology, 223

B

Baconian probabilities, 861
 Ball and beam system, 830
 BARON, 705
 Basically available, soft state and eventual consistency (BASE), 180
 Basic defuzzification distribution OWA (BADD-OWA) operator, 719
 Basic theory of numerical computing (BTNC), 694, 697, 700
 BAYESian test, 55
 Bayes nets, 865
 Bellman and Zadeh's fuzzy decision, 771
 Bernstein polynomials, 57
 Berz, Martin, 705
 Big data, 178, 192, 193, 195, 615, 616, 619, 621
 conflict resolution, 686–687
 definition, 671
 document security policy modeling and integration, 677–684
 GSP, 687–688
 local security policy sets, integration process for, 685–686
 security, 689
 UML extensions, for document modeling, 675–679
 use-case background, 673–675
 variety, 619–620
 velocity, 620
 veracity, 621
 volume, 619
 Binary Boolean set-operations, 325
 Binary chromosome, 828
 Binary continuous function, 565–567, 598–600
 Binary GrC model, 208
 Binary neighborhood system (BNS), 211, 224, 226, 228
 Binary normal probability density, 570
 Binary relation neighborhood system (BRNS), 225
 Binary sequence, 573
 Binary transaction data, 99
 Bioinformatics, 187
 accuracy comparisons, 273
 genetic neural support vector machines, 272
 hidden layers, 272
 protein structures, 272
 simulations, 273
 tertiary classifier, 272
 Biologically-inspired computing, 191

Biological neural network, 265, 266, 276
 Biometric identifier (BID), 642
 Bipolar de Morgan Brouwer-Zadeh distributive lattice, 141
 Bipolar Pawlak-Brouwer-Zadeh lattice, 139, 140, 142, 462, 463
 Black box problem, 266
 Boolean algebra, 501, 502, 508, 510, 514, 529, 532, 540
 of sets, 342, 349
 Boolean information base, 134, 135
 Boolean set-ring, 346
 Bootstrap technique, 907
 Borel- σ -algebra, 51
 Borel measurable function, 564
 Borel-measurable mapping, 901
 Boundary region, 878
 Boyce-Codd normal form, 184
 Brain informatics
 noise levels, 274–276
 pattern recognition, 274, 275
 simulation results, 275
 symmetry and similarity, 276
 Branch and bound algorithms, 701
 Breeder Genetic Algorithm (BGA), 823
 Brouwer negation, 463
 Brouwer-Zadeh lattice, 140
 Brute-force attacks, 634
 Buckley's derivation of Poss, 775
 Business logic/server tier, 234

C

Calculus, 340
 differential, 912
 integral, 912
 California Consumer Privacy Act (CCPA), 654
 Candidate information models
 attribute-based data model, 150
 extensible markup language, 152–153
 functional data model, 150–152
 future research, 158–159
 relational data model, 149–150
 resource description framework, 153
 suitability in support of granular computing, 155–158
 web ontology language, 153–155
 Cantor's set, 584
 Cartesian product, 325, 334, 343
 Case-based reasoning, 135–139
 Category, 327
 Category theory based GrC model, 209
 Cauchy-Schwarz inequality, 174
 Centralized log servers, 627
 Certainty systems, 584–585
 Chain-rule, 338
 Characteristic set, 395, 396
 Choice problems, 420
 Choquet integral, 721, 722
 Chromosome structure, 825
 Church's thesis, 511
 Clarke, B.L., 507

- Classical mereology, 504
 - Classical propositions, 386
 - Classical rough set, 120, 129, 130
 - vs. fuzzy set, 133–135
 - Classification knowledge, 518
 - Classification problems, 420
 - Classifier-based semantics, 2–4
 - Closure mereology (CM), 505
 - Cloud computing, 165, 166, 621
 - Clustering, *K*-means clustering, *see K*-means clustering
 - Cofinite, 329
 - Cohen, L.J., 861
 - Cointension, 34–35
 - Cointensive precisiation, 44
 - Collective attention spam, 247
 - Combination principle, 318
 - Combinatorial topology, 203
 - Command and control (C&C) centers, 635
 - Command and control (C&C) traffic, 627
 - Common vulnerability and exposures (CVE) entry, 633
 - Commutative group structure, 326
 - Commutative-ring structure, 327
 - Commutative semigroup structure, 326
 - Comparative possibility, 861
 - Complementation, 342
 - Completeness theorem, 534
 - Complete Pawlak theory (CPT), 163
 - Complete theory, 504
 - Comprehensibility, 113
 - Computation(s), 178, 180, 185, 192, 223
 - science, 911
 - Computational theory of perceptions (CTP), 45
 - Computational web intelligence
 - data parameters, 273
 - implementation, 273
 - input data, 273
 - performance, 274
 - Computer architecture, 911
 - Computing, 891–894
 - Computing with words (CW), 44, 181
 - Concept approximation, 215
 - Concurrency relation, 347
 - Concurrency control, 180
 - Conditional possibility, 863
 - Condition formulae, 884
 - Conditioning, 867–868
 - Confidential Information Protection and Statistical Efficiency Act (CIPSEA), 653
 - Connecticut Crash Data Repository (CCDR), 673
 - Consistent (deterministic), 884
 - Constraint propagation, 701
 - Constraint variable, 35
 - Contingency table
 - rank, 10–11
 - from rough sets, 8–9
 - Continuous functions, 335
 - Continuous fuzzy data set, 597
 - Continuous probability density, 565, 572, 573
 - Continuous stochastic systems, 565–578, 589
 - Conventional mathematical programming, 771
 - Conventional rule-induction methods, 382, 385
 - Cooperative multi-hierarchical query answering systems, 2–6
 - Cooperative query answering system, 4–5
 - Coordinate vectors, 497
 - Corresponding matrix, 9–16
 - COSY system, 705
 - Countable-additivity condition, 342
 - Couple, 325
 - ordered pair, 343
 - Coverage, 377, 378, 387
 - Crisp set, 768
 - Criteria, 421
 - Criterion, 420
 - Cross-Industry Standard Process for Data Mining (CRISP-DM), 252
 - Curse of dimensionality, 266, 268, 269
 - α -Cut, 898
 - CVD, 384
 - Cyber-attacks, 626
- D**
- Daily stock price dataset, 91
 - Data attribute descriptions, 260
 - Database management systems (DBMS), 180, 184
 - Data clustering, 728
 - Data confidentiality, 644
 - Data exfiltration, 636
 - Data mining (DM), 98–100, 178, 187, 190, 221, 223, 226, 356, 366, 370, 469, 480
 - association rule, 231, 232
 - attribute rule-formulae, 260
 - classification, 230
 - clustering, 231
 - definition, 252
 - dependency and granularity, 7–11
 - framework, 232
 - generalization operator, 258
 - generalization relation, 257
 - goal and application, 255
 - granularity, 256, 257
 - granular model, 253, 262
 - KDD process, 252
 - K-descriptive language, 259
 - knowledge description formulae, 260
 - knowledge rule-formulae, 260
 - knowledge system, 255, 256
 - K-satisfaction, 261
 - K-truth, 261
 - operators, 258
 - process, 258, 259
 - in rough databases, 368–369
 - semantic model, 253, 256
 - syntactic generalization, 252
 - syntax and semantics for, 253
 - Data owner approach, 687
 - Data privacy, 645
 - Data risk-partitioning process, 628

- Data set, 377
- D-core, 885
- Decidability, 501
 - effectively inseparable, 511
 - enumerable, 511
 - finitely inseparable, 511
 - first-order mereological theory, 512
 - LC, 512
 - procedure, 510
 - reasonable first-order language, 511
 - recursive, 511
 - recursively axiomatizable, 512
- Decision class, 424, 884
- Decision formulae, 884
- Decision making, 712, 713, 814
- Decision problems, 421
- Decision rule(s), 118, 122, 127, 130–132, 136, 138, 142, 280, 287, 421, 424–425, 439, 883–885
 - applications, 439–440
- Decision support, 421–423
- Decision systems, 280, 286–288, 883
- Decision table, 390, 391, 393–396, 424–425
- Decision-theoretic foundations, 865–866
- Decision under uncertainty, 420
- Declarative graduation, 27
- Decorations, 704
- Deep granular neural network, 265, 269, 270
- Definable set, 393
- Defuzzification, 768, 811
- Degree of granularity, 7, 15–16
 - and dependence, 15–16
- Degree of potential surprise, 860
- Delegation Diagram (DD), 677, 681–683
- De Morgan Brouwer-Zadeh lattice, 140
- De Morgan lattice, 539
- Dempster–Shafer combination rule, 318
- Dempster–Shafer theory, 181
- Dependency in data mining, 7–11
- Description logic, 279
- Descriptive datamining, 259, 260
- Descriptors, 884
- Desideratum, 34
- Determinantal divisors, 13
- Deterministic and non-deterministic decision tables, 523
- Dewey Decimal Classification (DDC), 494
- Diagonal matrix, 15, 16
- Differentiability, 338
- Dimensionality reduction, 885
- Dirac notations, 171–172
- Directed rounding, 701, 704
- Directory keywords, 150
- Discrete Interval Valued Type 2 Fuzzy Sets (DIVT2FS), 89
- Discrete stochastic systems, 578–581
- Discretionary access control (DAC), 616–618, 671–673, 675, 677, 680, 683, 686, 689, 690
- Discretionary access control model (DAC), 202
- Discretization, 390–392, 911
- Diverging to infinity, 331
- Divide-and-conquer, 222, 234, 268, 269
- Divisors and degree of dependence, 13–14
- Document-level security, 672
- Document Role-Slice Diagram (DRSD), 677, 679, 681–683
- Document Schema Class Diagram (DSCD), 673, 675–677, 683
- Document type definitions (DTDs), 152
- Domain-based message authentication, reporting and conformance (DMARC), 475
- Domain-flux, 635
- Domain generalization hierarchies (DGH), 657
- Domain generation algorithms (DGA), 635
- Domain keys identified mail (DKIM), 475
- Domain name system (DNS), 472
 - cache poisoning, 627
- Dominance, 420
 - relation, 438
- Dominance-based rough set approach (DRSA), 119, 121, 125, 423, 428–432
 - algebraic structure for, 139–142
 - ambiguity, 126
 - case-based reasoning, 135–139
 - complementarity property, 127
 - comprehensive weak preference relation, 125
 - condition attributes, 125
 - decision attributes, 125
 - decision rules, 122, 439
 - for decision under uncertainty, 445–450
 - dominance based constraints, 121
 - dominance cones, 121, 126
 - dominance principle, 126
 - downward union of classes, 125–127, 140
 - fundamental feature of, 125
 - fuzzy sets (*see* Fuzzy sets)
 - graded dominance relation, 438
 - inclusion properties, 127
 - information function, 125
 - multi-graded dominance relations, 440–441
 - multiple criteria choice and ranking context, 443–445
 - multiple criteria choice and ranking problems, 436
 - ordered data, 121
 - ordinal classification examples, 432–436
 - ordinal classification problems, 125
 - pairwise comparison table, 436–438
 - philosophical basis of, 123–125
 - ROR for, 436
 - upward union of classes, 125, 126, 140
 - VC-DRSA, 130–131
 - without degree of preference, 441–443
- Double-inputs and single-output continuous stochastic systems, 570–578
- Double-input single-output open-loop system, 572
- d*-query, 3–5
- Dwyer, P. S., 702
- Dynamic granule
 - big data, 193
 - time-varying data, 193
- Dynamic Host Configuration Protocol (DHCP) IP addresses, 628

E

- eClustering, 728
- t*-conorms, 712
- Electronic discovery (eDiscovery), 469, 477, 478
- Electronic mail (email), 469
- Elementary divisors, 14
- Elementary equivalence, 504
- Elementary set, 120, 393, 880, 881
- Elementary transformation, 14
- Email archives
 - acquire-store-retrieve, 470
 - clients, 476, 477
 - compliance/surveillance, 477
 - and data mining, 469
 - eDiscovery, 477, 478
 - email system, 472
 - format history, 470–472
 - implicit/hidden knowledge, 480, 483
 - indexing and querying, 480
 - mailboxes, 477
 - messages and attachments, 470
 - modes, 478
 - parsing, 478, 479
 - security and SPAM prevention, 473–476
 - storing, 479
 - tasks, 478
 - user association, 479
- Email clients, 476, 477
- Email format history
 - email headers, 471
 - envelope and content, 470
 - internal user names, 471
 - MIME, 471, 473
 - RFC 822, 471
- Email messages, 470
- Email security, 470
- Email segmentation, 480
- Email spoofing, 469, 473–476
- Email system, 472
- Embedded State, 210
- Emergence, 185–187
- Encoding-decoding mechanisms, 183
- Encryption, 642
- t*-norms, 712
- Ensemble learning, 399
- Entropy, 361–362
- Epistemic uncertainty, 369
- Equidistant partition data set, 597
- Equi-partition set, 336
- Equivalence class, 222, 326, 551–553, 879, 880
- Equivalence relation, 325, 346, 878
- Ershov's theorem, 512
- eSensor, 735
- Euclidean distance, 231, 488, 494
- Euclidean space, 551, 554
 - soft boundary in, 555
- Evanescence set of indices, 333
- Eventual set, 329
- Evolving Fuzzy systems (EFS)
 - adaptive real-time classifiers, 735
 - classification, 731–733
 - eClustering, 728
 - fault detection and prognostics, 736
 - Fuzzy logic controllers, 733, 734
 - intelligent or inferential sensors, 734
 - predictive models, 736, 737
 - signal reconstruction, 736
 - TS fuzzy systems, 728–731
- Evolving system, 725
- Exclusive rules, 379, 380
- Existence and uniqueness of solutions, 705
- Existing fuzzy-valued data, 904
- Expertise Power Graph, 480
- Exponential set, 344
- Extended interval arithmetic, 703–705
- Extended version of sequence, 335
- EXtensible markup language (XML), 152–153, 672, 673, 676, 677, 689, 690
- Extension, 327
 - of function, 330
 - of measure, 349
- Extensional closure mereology (CEM), 505
- Extensionality principle (EP), 505
- Extensional mereology (EM), 505, 507
- Extracted rules, 383
- Extrinsic uncertainty, 168

F

- Facebook, 240
- Facial image recognition, 488
- Falsity preserving property, 315
- Family of sets, 325
- Feasible alternative decisions, 767
- Federal Bureau of Investigation (FBI), 673
- Federal Trade Commission (FTC), 654
- Federated Crash Repository (FCR), 675, 677, 683–685, 687, 688, 690
- Fermat, 339
- Field of sets, 508
- Field structure, 326
- Fifth GrC model, 208
- Finite inseparability, 511
- Finitely axiomatizable, 504
- Finitely inseparable, 511
- Finite product (FP), 505
- Finite-rank, 353
- Finite structures, 509
- Finite sum (FS), 505
- First-order logic
 - assignment function, 503
 - atomic formula, 503
 - complete theory, 504
 - elementary equivalence, 504
 - formulas, 503
 - free and bound variables, 503
 - logical implication, 504

- First-order logic (*cont.*)
 - logical symbols, 502
 - logical theory, 504
 - parameters, 502
 - recursive axiomatizability, 504
 - satisfaction, 503, 504
 - sentence, 503
 - structure/interpretation, 503
 - symbols, 503
 - terms, 503
- FL-generalization, 43, 44
- Floating point arithmetic, 913
- Focusing mechanism, 376
- Footprint of uncertainty (FOU), 746
- Forward-wave-backward-wave learning
 - algorithm, 271
- Foundation of mathematics, 502
- Fourier and moments characteristics, 352, 353
- Fourier transformation, 173, 493
- Fourth GrC model, 208
- F-probability field, 784, 785
- Freiburger Intervallberichte*, 703
- Frequent set, 329
- Friend request, 240
- Function, 343
 - approximation property, of fuzzy systems, 598–601
 - computation model, 487
 - data model, 150–152, 155
- Fusion of knowledge, 281, 297–298
- Fuzzification
 - axioms, 513
 - conditions, 513
 - expressions, 513
 - first-order language, 513
 - GFM, 514
 - linear ordering, 514
 - principles, 512
 - semantics and syntactics, 513
- Fuzziness, 810
- Fuzzy approximation system, 569, 570
- Fuzzy arithmetic, 763
- Fuzzy banach spaces, 767
- Fuzzy binary granular structure, 228
- Fuzzy C-Regression model (FCRM), 81
- Fuzzy cluster, 84, 728
- Fuzzy conditional statement, 807, 808
- Fuzzy controller, 827
- Fuzzy control system, 827
- Fuzzy data mining, 101–103
- Fuzzy decision making, 768
- Fuzzy estimation, 901, 902, 906
- Fuzzy extension principles, 762, 764
- Fuzzy granular structure, 228
- Fuzzy graph constraint, 37
- Fuzzy-graph extension principle, 41
- Fuzzy implication(s), 845
 - operator, 595
- Fuzzy inequalities, 765
- Fuzzy inference reasoning, 845
- Fuzzy inference rules, 567, 568, 574, 575, 579–581, 588–590
- Fuzzy inference system, 846
- Fuzzy information base, 131
- Fuzzy information processing, 705
- Fuzzy integrals, 720–722
- Fuzzy intervals, 761, 762
- Fuzzy logic, 162, 177, 181, 191, 743, 812, 813
 - bivalent logic, 21
 - cointension, 34–35
 - cointensive precisiation, 31–33, 44
 - concept of granulation, 28
 - concepts of precisiation, 31–33
 - conceptual structure of, 22–24
 - constraint variable, 35
 - controllers, 733, 734
 - CTP, 45
 - deduction, 41, 42
 - eventual mechanization, 19
 - Fuzzy graph constraint, 37
 - GCL, 38–40
 - for generalization of scientific theories, 43
 - generalized constraint, 35
 - GPS system, 46
 - graduation and granulation, 20
 - granular vs. granule-valued distributions, 31
 - granulation and interpolation, 30
 - imprecise probabilities, 45
 - linguistic variables, 44
 - misconception, 21
 - modeling language, 45
 - possibilistic constraint, 36
 - precisiation and cointension, 44
 - primary constraints, 37
 - principal contributions of, 42
 - probabilistic constraint, 36
 - question-answering schema, 21
 - singular and granular values, 28
 - standard constraints, 37
 - usuality constraint, 37
 - veristic constraint, 37
- “Fuzzy max” concept, 778
- Fuzzy measure, 721, 722
- Fuzzy mereology, 513, 514
- Fuzzy neighborhood system (FBS), 229
- Fuzzy neural network (FNN), 268, 269
- Fuzzy neural network with knowledge discovery (FNNKD), 267
- Fuzzy number, 763, 809, 898–900, 905
- Fuzzy optimization
 - ambiguity, 765
 - Banach space, 767
 - classical approaches, 770
 - constraint coefficient matrix, 777
 - containment, 760
 - conventional mathematical programming, 771
 - degree of occurrence, 765
 - degree of satisfaction, 765
 - equality, 760

- extension principle, 762, 763
- flexible programming, 765
- fuzzy interval, 759, 761, 762
- fuzzy mathematical program, 769
- fuzzy max, 778
- fuzzy number, 759
- fuzzy objective function, 779
- fuzzy/possibilistic problems, 781
- fuzzy-valued function, 769
- inequality relationship, 782
- interactivity, 770
- intersection, 760
- IVPM set function, 783
- L-R fuzzy interval, 760
- Markowitz's mean-variance approach, 782
- mathematical analyzes, 756
- mathematical models, 756
- membership function, 774
- modal value, 759
- non-interactivity, 761
- nonlinear process, 756
- normative criterion, 757
- normative models, 757
- normative probability model, 757
- objective function, 776
- possibilistic decision making, 769
- possibilistic linear program, 775
- possibilistic mathematical program, 780
- possibilistic programming, 775
- possibilistic uncertainty, 755, 756
- possibility distributions, 765
- possibility theory, 764, 766
- probability distribution, 784, 787
- real Euclidean spaces, 768
- real-valued function, 762
- relation, 760, 766
- surprise function, 773, 774
- symmetric triangular fuzzy number, 760, 776
- trapezoidal fuzzy interval, 760
- triangular fuzzy number, 760
- Zimmermann, 772
- Fuzzy pairwise information base, 137, 138
- Fuzzy probability theory
 - analysis of sensitivities, 66
 - application fields, 65–67
 - definition, 52
 - design and decision making, 67
 - future research, 67–68
 - fuzzy random variables, 58–60
 - mathematical environment, 53–58
 - model validation and verification, 67
 - reliability assessment, 66
 - representation of fuzzy random variables, 63–65
 - subjective assessment of temperature, 52
- Fuzzy programming, 771
- Fuzzy random event, 60
- Fuzzy randomness, 54
- Fuzzy random sets, 38
- Fuzzy random variable, 899–901, 904–907
- Fuzzy reasoning, 844, 849
- Fuzzy relations, 760, 766, 770, 804–806
- Fuzzy rough relational database, 363
 - entity-relationship diagram, 364
 - fuzzy rough difference, 364
 - fuzzy rough intersection, 365
 - fuzzy rough join, 366
 - fuzzy rough projection, 366
 - fuzzy rough selection, 365
 - fuzzy rough tuple, 363
 - fuzzy rough union, 365
 - interpretation, 363
- Fuzzy Rule-based (FRB) system, 727
- Fuzzy Rule Base Models, 92
- Fuzzy rules, 831
- Fuzzy sample mean, 906
- Fuzzy sets, 51, 119, 127, 165, 177, 180, 487, 717, 727, 743, 744, 793–796, 891–893, 897, 899, 900, 903
 - approximation, 229, 230
 - complementarity property, 129
 - defuzzification, 811
 - dominance-based rough approximation, 131–133
 - dominance relation, 128
 - downward union, 129
 - extension principle, 808, 809
 - fuzzy number, 809
 - idea, 795
 - inclusion property, 129
 - information granularity, 228, 229
 - monotonic rough approximation, 133–135
 - multistage decision making, 816, 817
 - operations, 802–804
 - and possibility theory, 759
 - probabilities, 810
 - properties, 797–800, 802
 - upward union, 128
 - weak preference relation, 128
 - Zadeh's, 796, 797
- Fuzzy set theory, 100, 101, 181, 183, 221, 222, 360, 755
 - characteristic functions, 356
 - concentration operation, 357
 - dilation operation, 357
 - support and α -cuts, 356
- Fuzzy similarity for parallel function computation
 - model, 487
- Fuzzy statistical tests, 902, 903
- Fuzzy system-based intelligent systems, 737
- Fuzzy system modeling (FSM)
 - daily stock price dataset, 91
 - desulphurization facility, 91
 - DIVT2FS, 89
 - income prediction model, 91
 - Lagrange function, 86
 - OLSE method, 82–84
 - secondary membership function, 87, 88
 - SFF-LSE model, 87
 - SFFM, 92, 93
 - special fuzzy functions, 81, 82, 90
 - structure identification methods, 89

- Fuzzy system modeling (FSM) (*cont.*)
 - SVM for regression, 84, 85
 - SY-FRB, 92
 - TS-FRB structures, 80, 92
 - type 2 fuzzy sets, 78
- Fuzzy system representations, of stochastic systems
 - discrete stochastic systems, 578–581
 - double-inputs and single-output continuous stochastic systems, 570–578
 - reducibility, 581–584
 - uncertainty systems with one dimensional random variable and representations, 584–590
 - unification on uncertainty systems, 590
- Fuzzy systems, 593, 712, 722, 794
 - approximate remainder formula of $f_n(x)$ to continuous function $s(x)$, 601–604
 - error estimation, fuzzy system $S_n(x)$ and $f_n(x)$, 605–611
 - function approximation property, 597–601
 - simulation, 612–613
 - structures of, 593–597, 725
- Fuzzy test, 902
- Fuzzy uncertainty, 758
- Fuzzy-valued function, 769
- Fuzzy vector, 51

- G**
- Gauss, Carl Friedrich, 702
- Gaussian functions, 848
- Gaussian membership functions, 847–849, 851
- Gaussian radial base kernel, 86
- General Data Protection Regulation (GDPR), 483
- General extensional mereology (GEM), 502, 505, 506
- General extensional mereotopology (GEMT), 507
- Generalizations, and rough approximation, 551–553
- Generalized constraint, 35, 36
- Generalized constraint language (GCL), 38–40, 42
- Generalized decision class, 884
- Generalized decision function, 884
- Generalized information theory, 53
- Generalized OWA operator, 719
- Generalized rough sets, 546–547
- Generalized theory of uncertainty (GTU), 45
- Generalized type-2 fuzzy sets, 747
- Genetic algorithms (GAs), 103, 191, 822–824, 828
 - approach, 822
 - crossover operation, 103
 - mutation operation, 103
 - selection operation, 103
- Genetic cycle, 826
- Genetic-fuzzy data mining techniques, 104, 105
 - DGFSMS problem, 110–112
 - IGFMMS problem, 109
 - IGFSMS problem, 105–108
- Genetic granular neural network (GGNN), 273
- Geometric mean, 716, 722
- Geometric OWA (GOWA) operator, 719
- Gini index, 370
- Global backward-wave learning algorithm, 271
- Global optimization, 701
- Global security policy (GSP), 683, 686–688
- Gödel number of α , 511
- Goodman, N., 506
- Graded modal logic (GML), 412
- Gradual decision rules, 122, 142
- Gradual number, 761
- Graduated granulation, 20
- Granular computing (GrC), 163, 166, 177, 221, 539–540
 - advantages, 402
 - AI-engineering, 192
 - Ancient Example, 205
 - applications, 235
 - basic components of, 119
 - big data, 195
 - candidate information models (*see* Candidate information models)
 - category theory based GrC model, 209
 - category theory based models, 209
 - cognitive informatics, 194
 - computational models, 192
 - computer security, 204
 - concept approximation, 216
 - in database theory, 164
 - data mining, 189–190, 230–232
 - definition, 147–148, 182, 201, 401
 - and DRSA (*see* Dominance-based rough set approach (DRSA))
 - dual of granulation, 212
 - early GrC models, 210
 - expressive power, 156–157
 - extensibility and utility, 157–158
 - formalisms, 178
 - formalism and theory, 155–156
 - formalization, 200
 - formal models of, 205
 - foundations of, 235
 - framework, 235
 - fuzzy set system, 228–230
 - and fuzzy set theory, 165
 - general issues in, 165
 - granular and related structures, 210
 - granular structure (GrS), 202
 - human body, 204
 - human society, 205
 - information granules, 192–194
 - information hiding and emergence, 185–187
 - information structure, 213
 - integrations on partitions, 213
 - intelligent problem solving, 187–189
 - issues, 179
 - knowledge structure, 202
 - linguistic structure, 202
 - local granules of uncertainty, 204
 - local GrC model, 206
 - methodological and algorithmic issues, 195
 - methodology, 402
 - non-homogeneous methods, 183
 - non-partition case, 214

- non-partition theories, 222
- NS, 224–226
- and ontologies, 185
- partitioning, 181, 204
- periphery of, 194
- perspectives, 223
- philosophical foundation, 178, 179, 195
- pre-GrC historical notes, 207
- questions and observations, 542–543
- quotient Structure (QS), 202
- realizes modern example, 207
- rough set system, 226–228
- S5 rough algebra, 540–542
- second GrC model, 203
- simple partition, 212
- simplicial complex, 204
- in social networks, 164–165
- soft and natural computing, 190, 191
- software engineering, 232–235
- space and time, 204
- straightforward application, 415
- suitability of information models for, 155–158
- terminology, 222
- two granular and quotient structures, 212
- Granular data, 265, 267
- Granular data model (GDM), 164
- Granularity, 177, 180, 182–185, 193, 222, 255–257, 312
 - in data mining, 7–11
 - relation, 314
- Granular learning algorithms
 - cognitive neural granules, 270
 - forward-wave-backward-wave learning, 271
 - global backward-wave learning, 271
 - local forward-wave learning, 270
- Granular link, 265
- Granular mathematics (GrM), 203
- Granular model, 253, 262
- Granular neural network, 265
 - bioinformatics, 271–273
 - brain informatics, 274–276
 - computational web intelligence, 273–274
 - data fusion, 265
 - data granulation, 267
 - divide-and-conquer strategy, 268
 - FNN, 268
 - intelligent techniques, 276
 - learning algorithms, 270–271
 - linguistic extraction layer, 267
 - multi-FNNKD layer, 267
 - multimedia granules, 267
 - output layer, 268
 - soft computing system, 268
- Granular neuron, 265
- Granular probability distribution, 36
- Granular structure (GrS), 177, 184–185, 202, 211
- Granular topology, 697
- Granular weights, 265
- Granular world, 178
- Granulation, 29, 177, 180–182, 187–189, 192, 223
 - definition, 402
 - functional, 402, 404, 407–408
 - relational, 402, 404, 408–410
 - from relational to functional, 413–415
 - types of, 404
- Granulation of knowledge
 - binary general relations, 283
 - calculus on granules, 293–294
 - in data mining, 299–307
 - definition, 281
 - generalized descriptors, 284
 - granules based on indiscernibility in information systems, 283
 - purpose of, 282
 - reasoning about uncertainty, 294–297
 - reasoning in cognitive schemes, 298–299
 - reasoning in distributed system, 297–298
 - rough inclusions, 284–285, 289–293
 - similarity, 289
- Granule, 177, 180, 181, 204, 224, 267
 - vs. attribute/entity (database modeling), 182
 - vs. cluster (data analysis/data mining), 182
 - measurement, 183–184
 - vs. OO paradigm, 182
 - vs. system vs. graph, 182
 - vs. word, 181
- Gravity method, 596
- Green's theorem, 183
- Grid computing, 165
- Ground fuzzy mereology (GFM), 514
- Ground mereology (GM), 505, 508, 511
- Group, 344
- Group-homomorphism, 345
- Guaranteed possibility, 862, 865, 867
- Gunk, 506
- H**
- Hansen, Eldon, 702, 703
- Harmonic mean, 716, 722
- HAUSDORFF metric, 56
- Headache
 - negative rules, 383
 - positive rules, 383
- Heine–Borel theorem, 183
- Hermitean operator, 169, 170
- Heuristic, 911
- Hierarchical approach, 686, 687
- Hierarchical genetic algorithm (HGA), 822
 - approach, 822
 - method, 831
- Hierarchy, 177, 182–186
- Homeomorphism, 328, 345
- Human-centered information processing, 180, 192, 194, 195
- Human cognition, 317
- Human information processing system (HIPS), 194
- Hybrid logics (HL), 412, 413
- Hybrid neural networks, 266, 269

- Hybrid soft computing
 - anesthesia control, 826, 827
 - ANFIS model, 834
 - automatic anesthesia control, 822
 - bar and ball system, 830
 - different methods, 833
 - evolution of fuzzy systems, 824–826
 - fuzzy controller, 827, 829
 - genetic algorithm, 822–824, 828
 - genetic algorithms for NN, 832–834
 - HGA, 829
 - magnification, 833
 - neuro-genetic model, 835
 - simulation results, 832
 - time series prediction, 834
 - type-1 fuzzy models, 835
 - type-2 fuzzy model, 834
 - Hyperfinite coin-tossing, 351
 - Hyperfinite equi-partition, 336
 - of the unit-interval, 351
 - Hyperfinite random walk, 352
 - Hyperfinite-rank, 353
 - Hyperfinite set, 336
 - Hyperfinite time-axis, 351
 - Hyperinteger, 334
- I**
- Idealization principle, 347
 - Identity of indiscernibles, 123–125
 - IEEE 1788–2015, 704
 - IEEE 754–2019, 704
 - IEEE Computational Intelligent Society, 223
 - Implication algebra, 536
 - Imprecise data, 897, 898, 901, 904, 907, 908
 - Imprecise perceptions, 897
 - Imprecise probabilities, 45
 - Imprecision, 745
 - Income prediction model, 91
 - Inconsistent data, 393–394
 - Inconsistent (non-deterministic), 884
 - Independent increment (Hyperfinite-Rank), 352
 - Indirect learning-based control scheme, 734
 - Indiscernibility-based rough set approach (IRSA), 121, 122
 - Indiscernibility relation, 518, 878–881
 - Induced OWA (IOWA) operator, 719
 - Infimum, 328
 - Infinitely close, 333
 - Infinitesimal, 329, 331, 333
 - difference, 324, 333
 - Infinites, 328, 329, 331, 333
 - Information fusion, 711–714
 - Information granularity, 228, 229
 - Information granule, 180, 192–194, 879, 881, 885
 - calculi, 881
 - Information hiding, 184–187, 190, 215
 - Information integration, 187, 215, 235, 711, 713, 714, 723
 - Information retrieval systems (IR), 180
 - Information system (IS), 280, 286, 882, 883, 885
 - indiscernibility relation, 227
 - machine learning, 227
 - reduct, 227
 - Information table (IT), 164, 882
 - Information theory, 360–361
 - Infrastructure as a service (IaaS), 621
 - Injective homomorphism, 345
 - Insider threats, 636
 - Instant messages (IM), 470
 - Intelligent or inferential sensors, 734
 - Intelligent problem solving
 - construction under constraints, 188
 - granulations characterization, 188
 - GrC, 188–189
 - Interactive Multiobjective Optimization using Dominance-based Rough Set Approach (IMO-DRSA)
 - characteristics, 457–459
 - for financial portfolio, 459, 460
 - multiobjective optimization procedure, 454
 - for project portfolio, 460–462
 - steps, 452, 453
 - Interestingness, 113
 - Interesting rules, 383
 - Interlocking hierarchy, 186
 - Intermediate-value, 337
 - Internal set theory, 347, 353
 - Internet Message Access Protocol (IMAP), 476
 - Interpolation function, 597
 - Interval arithmetic
 - applications, 706
 - definition, 701
 - development and use of, 702
 - early history, 702
 - future directions, 706
 - general and introductory books, 705
 - modern history, 702
 - resources, 706
 - software packages, 706
 - uses, 705
 - Interval computing, 701
 - Interval Newton methods, 705
 - Interval type-2 fuzzy logic system (IT2 FLS), 750, 751
 - Interval type-2 fuzzy sets (IT2 FSs), 747–749
 - Interval-valued fuzzy sets, 797, 863
 - Interval-Valued type 2 fuzzy sets (IVT2FS), 79, 88
 - INTLAB, 706
 - Intrinsic uncertainty, 168
 - Intrusion detection system events, 627, 634
 - Intrusions, 634
 - Invertible function, 717
 - Irreducible implicative filter, 537
 - Isolated State, 210
 - IVPM set function, 783

J

Jaulin, Luc, 705
 Joint probability space, 564, 572, 581
 Journal, Reliable Computing, 706
 Journaling, 478

K

k-anonymity model, 656
 Karush–Kuhn–Tucker (KKT) conditions, 768
 Karush–Kuhn–Tucker (KKT) theorem, 87
 Kaucher arithmetic, 705
 K-descriptive language, 259
 Kepler conjecture, 705
 Ket vector, 171
 Klein’s Erlangen program, 200
 Knowledge, 279

- attribute descriptions, 260
- decision systems, 287–288
- engineering, 215
- fusion of, 281
- granulation (*see* Granulation of knowledge)
- information systems, 286–287
- representation, 164, 215, 279, 523
- system, 254, 255

 Knowledge discovery in databases (KDD), 98, 190
 Knowledge structures (KS), 202, 211
 Kolmogorov probability space, 342
 Kolmogorov’s axiomatization, 342
 Kronecker property, 565, 569, 576, 580, 597
 Kwakernaak/Kruse and Meyer’s sense, 900

L

Lagrange multiplier method, 86
 Large Sample Theory, 906
 Learning from Examples based on Rough Sets (LERS), 390, 391, 393, 397, 398
 Learning from Examples Module, version 2 (LEM2) algorithm, 390, 392
 Least squares method, 564
 Leaving-one-out method, 398
 Lebesgue algebra, 351
 Legal hold, 469
 Leibniz’ product rule, 339
 Leibniz program, 328
 Leibniz’s law, 123
 Leonard, H.S., 506
 Leśniewski, S., 501
 Leśniewski’s theory, 501, 502, 506
 α -level set, 51, 898
 Lewis, D., 861
 L-fuzzy sets, 23
 Lifting of function, 350
 Limited hyperreals, 333
 Limit object, 137
 Lindenbaum–Tarski algebra, 530, 532, 533, 541

Linear algebra, 7, 9, 14, 15
 Linear interpolation function, 597
 Linear spectral pairs (LSP), 737
 Linguistic structure, 202
 Local complementation (LC), 512
 Local forward-wave learning algorithm, 270
 Lockheed Missiles and Space Company, 702
 Loeb construction, 349
 Loeb content, 350
 Loeb integral, 350
 Loeb measure, 350
 Logical implication, 504
 Logical theory, 504
 Logical-type neuro-fuzzy systems, 847
 Logical-type reasoning method, 844
 Lorenz equations, 705
 Losonczy’s means, 718
 Lower approximation, 421, 878, 879, 881, 884
 Lower membership function (LMF), 748
L-R fuzzy interval, 760
L-R fuzzy numbers, 778
 LT-fuzzy sets, 545–546
 Lukasiewicz-type S-implication, 847

M

Machine learning, 227, 230, 232
 Mackey–Glass time series, 839, 840
 MAC Secure Information Diagram (MSID), 677, 679, 681–683
 Mailbox crawling, 478
 Mail transfer agent (MTA), 472
 Makino, Kyoko, 705
 Malicious domains, 635
 Malware propagation, 242
 Mamdani approach, 849
 Mamdani fuzzy model, 840
 Mamdanirule, 22
 Mamdani-type neuro-fuzzy system, 848
 Mamdani-type reasoning method, 844
 Mandatory access control (MAC), 616, 618, 671–673, 675, 677, 679, 683, 686, 689, 690
 MapReduce, 192, 193
 Marginal distributions, 9
 Markowitz’s mean-variance approach, 782
 Master Classification Index, 684
 Master Delegation Index, 685
 Matching rule, 398
 Mathematica, 706
 Mathematical modeling, 756
 Matrix algebra, 16
 Matrix attachment regions (MARs), 188
K-means clustering, 555–556

- allocation rule, 557
- K*-means algorithm, 556–557
- optimization problem, 559–560
- rough approximations, 557–559

- Measurable functions, 350
 - Membership function, 595
 - Meningitis, positive rules, 383
 - Mereological axioms, 504
 - atomicity, 505
 - atomlessness, 505
 - complementation, 505
 - greatest member, 505
 - overlap, 504
 - proper part, 504
 - UF, 505
 - underlap, 504
 - Mereological theory, 504
 - of concepts, 289
 - GEM, 505, 506
 - GM, 505
 - orderings, 505
 - Mereology, 281
 - being connected to, 507
 - Boolean algebras, 515
 - contact, 507
 - first-order logic, 502–504
 - first-order mereological axioms and theories, 504–506
 - formulations, 501, 506–507
 - fuzzification, 512–514
 - internal relation, 515
 - mathematical theory, 502
 - undecidable, 502
 - Mereotopology, 507, 515
 - Meta-problem, 894
 - Meterstick, 696
 - Min-based conditioning, 867
 - Minimal closure mereology (CMM), 505
 - Minimal extensional mereology (MEM), 505
 - Minimal length principle, 885
 - Minimal mereology (MM), 505
 - Missing attribute values, 390, 394–395
 - lower and upper approximations, 395–396
 - probabilistic approximations, 396–397
 - Mixed structures, 509
 - MMUCC, 675, 677–682, 690
 - Modal logic, 404, 410–411
 - graded, 412
 - propositional, 411–412
 - Modal mereology, 515
 - Models of mereological theories
 - anti-symmetry, 508
 - atomless models, 510
 - decidability, 510–512
 - foregoing analyses, 510
 - infinite models, 510
 - logical theory, 508
 - mixed models, 510
 - mutually disjoint subclasses, 509
 - Th(K), 509
 - upper semilattice, 508
 - Model validation and verification, 67
 - Moderate numbers, 330
 - Modified LEM2 (MLEM2) algorithm, 390, 397
 - Monad of infinitesimals, 334
 - Monotonicity, 119, 122, 125, 132, 133, 136
 - theorem, 525
 - Monotonic relationships, 121, 122, 125
 - Monte Carlo simulation, 57
 - Moore, Ramon, 702, 704
 - Multi-granular computing
 - average precision, 321
 - combination of attribute functions, 319
 - combination of QuotientSpaces, 319
 - combination of universes, 319
 - computational complexity, 315–317
 - falsity preserving property, 315
 - granularity relation, 314
 - hierarchy, 315
 - initialization, 320
 - mean and variance matrices, 320
 - multi-granular spatio-temporal objects, 312
 - quotient-space model, 318
 - truth preserving property, 314
 - Multi-granular structure, 311
 - Multi-hierarchical decision system, 2–4
 - Multi-input multi-output (MIMO) rule, 751
 - Multi-objective genetic-fuzzy data mining, 113
 - Multiple attribute/multiple criteria decision problems, 420
 - Multiple criteria decision analysis, 451–452
 - Multiple-minimum-support fuzzy-mining (MSFM), 99, 113
 - Multiple Objective Linear Programming (MOLP)
 - problem, 454
 - Multipurpose internet mail extensions (MIME), 469, 471
 - Multistage decision making, 816
 - Multi-valued representation/granulation, 229
 - MySpace, 242
- N**
- Naïve Bayes algorithm, 474
 - Naive set operations, 343
 - Natural computing, 191
 - Natural language processing (NLP)
 - applications, 493
 - book cataloging, 493–497
 - mathematical transformations, 493
 - translation, 497–498
 - word2vec, 493
 - Necessity of Strict Dominance index (NSD), 903
 - Negative infinity, 328
 - Negative rules, 378–380
 - CVD, 384
 - headache, 383
 - Neighborhood, 551, 553–555, 558, 561
 - Neighborhood system (NS), 177, 181, 221, 222
 - approximation, 225, 226
 - non-empty subsets, 224
 - notions, 224
 - Network flow data, 627
 - Neumaier, Arnold, 706
 - Neural net, 487, 488

- Neural network (NN), 832–834, 843
 - approach, 822
 - model, 834
- Neuro-fuzzy systems (NFS), 727, 893
 - Akaike estimation, 854
 - automobile transmissions, 844
 - camcorder focusing, 843
 - criteria isolines, 854
 - criteria isolines method, 853
 - fuzzy inference engine, 846
 - fuzzy inference reasoning, 845
 - fuzzy inference system, 846
 - fuzzy reasoning, 844
 - Gaussian membership functions, 847, 848
 - learning process, 853
 - logical-type neuro-fuzzy systems, 847
 - mamdani-type neuro-fuzzy systems, 848
 - pattern classification, 852
 - polymerization problem, 855
 - recursive procedure, 853
 - simplified neuro-fuzzy systems, 849, 850
 - subway operations, 844
 - Takagi–Sugeno-type model, 850, 851
 - TV color tuning, 843
 - washing-machine automation, 843
 - with weights, 851
- Newton, 912
- Newton method, interval, 705
- Ninth GrC model, 209
- NL-computation, 44
- Non-archimedean infinitesimal, 329
- Non-directory keywords, 150
- Non-free ultra filter, 332
- Nonmonotonic inference, 863–864
- Non-probability-qualified granule, 229
- Nonzero infinitesimal, 338
- Normalized linear, 802
- Normed counting measure, 351
- NoSQL databases, 621
- NS calculus, 338
- NS integral calculus, 340
- Nullifying sequence, 329, 331
- Numerical algorithms, 911
- Numerical analysis, 911, 912
- Numerical computations, 891
- O**
- Object-oriented (OO) paradigm, 182
- Object-oriented system design, 234
- On-line analytical mining (OLAM), 190
- On-line analytical process (OLAP), 190
- Online interactive map, 185
- Ontology, 6, 185, 497
- Open computing language (OpenCL), 489–492
 - execution model, 489
- Operational workflow, 480
- Order-completeness, 348
- Ordered-field, 327
 - extension, 326
- Ordered knowledge base, 141
- Ordered weighted averaging (OWA) operator, 718, 719
- Ordering, 331
- Ordinal classification, 421
- Ordinal properties, 118, 120, 121, 125, 136, 138, 142
- Ordinary differential equations, 912, 913
- Ordinary LSE (OLSE) method, 82–84
- Organization power graph, 480
- Overflow and underflow, 348
- P**
- Pairwise comparison table (PCT), 436–438
- Parallel computation, 911
- Parallel computing
 - NLP, 493–498
 - OpenCL, 489–492
- Parametric dimensionality, 268
- Pareto optimal solutions, 454, 713
- Parthood relation, 501, 504, 512
- Partial ordering, 501, 505, 508, 514
- Particle-wave duality, 168
- Partitioning approach, 186, 629
- Pawlak–Brouwer–Zadeh lattice, 140
- Pawlak rough set theory
 - approximation space, 518–519
 - set approximation regions, 519–520
- Pawlak’s model, 253
- Peirce algebras, 536–537
- P-elementary sets, 423
- Perfileeva transform, 30
- Permutation-based anonymization process, 660
- Personally identifiable information (PII), 483
- Philosophy of science, 178–180
- Phishing, 246, 473
- Platform as a service (Paas), 621, 622
- Point-set mapping, 595
- Policy integration, 673, 682, 687, 689
- Policy-level security, 672
- Polynomial kernel, 86
- Positional equivalence
 - definition of, 404
 - different notions of, 404
 - logical characterization of, 413
 - practical computation of, 416
 - relational granulation-based, 416
- Positive implication algebra, 535
- Positive infinity, 328
- Positive rules, 378
 - CVD, 384
 - headache, 383
 - meningitis, 383
- Possibilistic constraint, 36
- Possibilistic decision making, 769
- Possibilistic logic, 864–865
- Possibilistic mathematical program, 780
- Possibilistic uncertainty, 369, 370, 756
- Possibility distribution, 861–865, 867, 868

Possibility theory, 45, 755, 764
 applications, 871–872
 Cohen, L.J., 861
 definition, 859–860
 future directions, 871–872
 historical background, 860
 Lewis, D., 861
 qualitative (*see* Qualitative possibility theory)
 quantitative (*see* Quantitative possibility theory)
 Shackle, G.L.S., 860–861
 Zadeh, L.A., 861

Post Office Protocol Version 3 (POP3), 476

Power mean, 716, 717

Power-set, 344

Precisiated Natural Language (PNL), 44

Precision, 570

Preference model, 420

Pre-robinson Infinitesimals, 325

Primary constraints, 37

Prime implicative filter, 537

PRIMEROSE-REX, 380, 381

Privacy, 617–618

Privacy enhancing techniques
 biometrics authentication, 641
 data mining, 641
 record matching, 640

Privacy preservation
 anonymity, 653
 in big data analytics, 653
 big data mining issues, 666
 big data publishing issues, 664
 challenges on Big Data, 662
 data mining, 662
 data publishing, 654–656
 issues, 650
 NP-hard problem, 663
 open issues for, 664
 in re-identification problems, 655
 techniques for, 662
 traditional methods for, 664

Privacy preserving data mining (PPDM), 653

Privacy preserving data publishing (PPDP), 653

Probabilistic constraint, 36

Probabilistic decision tables, 523
 attributes, 523
 dependencies in, 524–526
 deterministic and non-deterministic, 523
 elementary sets, 524

Probabilistic rules, 377, 378

Probabilistic uncertainty, 369, 370

Probability content and measure, 349

Probability density, 168, 581
 function, 585, 588

Probability distribution functions, 55

Probability-possibility transformations
 mean values in possibility theory, 870–871
 possibility distributions induced by prediction intervals, 869–870
 from possibility to probability, 868–869
 subjective probability distribution, 870

Probability space, 342, 585, 586, 588

Product probability measure, 572

Proper parts principle (PPP), 505

Propositional modal logic (PML), 412

Q

Quadratic mean, 716

Qualitative possibility theory, 863
 decision-theoretic foundations, 865–866
 nonmonotonic inference, 863–864
 possibilistic logic, 864–865

Quality of classification, 424

Quantitative possibility theory
 conditioning, 867–868
 definition, 866–867
 possibility as upper probability, 867

Quantum computing, 191

Quantum postulates, 169–171

Quantum uncertainty, 173–174

Quasi-arithmetic means, 716–717

Quasi-Boolean algebra, 539

Quasi-Brouwer-Zadeh distributive lattice, 140

Quasi-Identifiers (QI) attributes, 656

Quasi-OWA (QOWA) operator, 719

Quasi-weighted mean, 717

Query Answering System (QAS), 1, 2, 5, 6

Query entropy (Q-entropy), 189

Quoted-Printable encoding, 472

Quotient algebra, 327

Quotient by eventuality, 331

Quotient set, 326, 880

Quotient-space model, 228, 316, 318

Quotient State, 211

Quotient structure (QS), 202, 211, 313

R

Random fuzzy set, 900

Random sets, 900

Random vector, 590

Rank
 of contingency table (2×2), 9
 of contingency table ($m \times n$), 10–11

Rank and degree of dependence
 degree of granularity and dependence, 15–16
 determinantal divisors, 13
 divisors and degree of dependence, 13–14
 elementary divisors and elementary transformation, 14
 and subdeterminant, 11–12
 subdeterminants and degree of dependence, 14
 submatrix, 11–12

(Rank j) Random Walker, 352

Rational fraction, 602

Rational monotony, 864

Reasoning, 281

Record matching, 640

Recursive least squares (RLS) method, 731

- Recursively axiomatizable, 504
 - Reduct property, 286
 - Reference object, 137
 - Relational database, 164, 180
 - Relational data model (RDM), 149–150
 - Relative information content (RIC), 184
 - Reliability assessment, 66
 - Reliable computing journal, 706
 - Remote Procedure Call (RPC), 477
 - Representation theorem (RT), 749
 - Resource description framework, 153
 - RFC 822, 471
 - Richardson, 913
 - Riemann-integrable function, 340
 - Riemann integral, 341
 - Riemann sum, 565–567, 573, 574, 576, 580, 590, 596, 598, 599
 - Risk-partitioning process, 629
 - Robotics, 705, 712
 - Robust Ordinal Regression (ROR), 436
 - Role-based access control (RBAC), 616–619, 622, 671–673, 675, 677, 680, 682, 683, 685–690
 - Rolle, 339
 - Rolle differential intermediate value theorem, 603
 - Rough approximation
 - and generalizations, 551–553
 - K*-means clustering, 557–559
 - and set-valued mapping, 553–555
 - Rough entropy, 361
 - Rough equality, 530, 540, 547
 - Rough filter, 545
 - Rough inclusions
 - definition, 281
 - from functors of many-valued logics, 291
 - granulation by, 289–293
 - granules from, 284–285
 - granules from parameterized variants of, 301–307
 - information system, 290
 - from metrics, 290–291
 - tolerance relations, 285
 - transitivity, 291
 - Rough mereological approach, 880
 - Rough relational database, 359
 - and entropy, 361–362
 - features, 359
 - Rough relation entropy, 362
 - Rough schema entropy, 362
 - Rough set(s), 120, 279, 358, 376, 420, 881
 - decision table, 120
 - fuzzy, 358
 - information table, 120
 - lower approximation, 120
 - spatial data, 366–368
 - upper approximation, 120
 - Rough set approach
 - approximation, 423–424
 - data table and indiscernibility relation, 424–425
 - decision table and rules, 424–425
 - dependence and reduction of attributes, 424
 - traffic signs example, 425–428
 - Rough set theory (RST), 163, 164, 177, 181, 183, 192, 221, 222, 402, 403, 407, 408, 416, 421–423, 529, 696, 893
 - approximation, 226
 - information system, 226–228
 - partition, 226
 - quotient space, 228
 - Roundoff error analysis, 701, 911, 912
 - backward, 913
 - forward, 913
 - R-probability field, 785
 - Rule-based headache information organizing system (RHINOS), 384
 - Rule-based system for headache (RH), 385
 - Rule induction, 380, 390–392, 397–399
 - methods, 385
 - Rump, Siegfried, 703
 - Russell's paradox, 501, 502
- ## S
- Safe and lazy approach, 686, 687
 - Sample decision system, 883
 - Samy worm, 242
 - Satisfaction levels, 814
 - Saturation principle, 348
 - SCAN meetings, 703
 - Scientific computing, 911, 913
 - Secure information diagram (SID), 677, 679, 680, 683
 - Secure multiparty computation (SMC) techniques, 640
 - Secure socket layer (SSL) vulnerability, 632
 - Security assurance, 671, 675, 685, 689, 690
 - Security filtering, 476
 - Security Information and Event Management (SIEM)
 - products, 626
 - Security modeling, 673, 689
 - Security-relevant data, 626
 - Security situational awareness systems (SSAS), 626
 - Semantic(s), 2–5, 185, 529
 - model, 257
 - Semi-supervised learning, 399
 - Sender policy framework (SPF), 475
 - Sensitive attributes (*SA*), 656
 - Separation of duty (SOD), 680
 - Sequence, 325
 - spaces, 330
 - Set of coral pores, 881
 - Set-set mapping, 595
 - Set-theory, 341, 502
 - Set-valued mapping and rough approximations, 553–554
 - Euclidean space, soft boundary in, 555
 - neighborhood, 554–555
 - Seventh GrC model, 209
 - SFF-LSE model, 82, 84
 - SFF-SVM models, 84
 - Shackle, G.L.S., 860–861
 - Shafer plausibility function, 866
 - Shannon entropy, 369
 - Short message service (SMS), 470
 - Signed infinities, 328

- Signum function, 339
 - Similarity, 280, 496
 - Simple Mail Transport Protocol (SMTP), 469, 470, 472
 - Simplified neuro-fuzzy systems, 849, 850, 852
 - Simulation, 911
 - Single-input single-output open-loop system, 564
 - Single learning systems, 856
 - Single-minimum-support fuzzy-mining (SSFM), 99, 113
 - SISO static open-loop system, 594
 - Sixth GrC model, 209
 - Smaller infinitesimal, 334
 - SMV algorithm, 86
 - Social network
 - data, 660
 - definition, 408
 - theory and applications of, 415
 - Social network analysis (SNA), 402
 - Social networking communities, 239
 - attacks against, 247–249
 - background on, 240–241
 - campaigns, 248
 - consequence of, 245
 - damage, 242
 - misinformation, 249
 - and privacy, 240
 - rapid worm propagation in, 242
 - social spam, 247
 - success of, 240
 - Social spam, 247
 - Soft computing, 190, 711, 892–894
 - Software as a service (SaaS), 621, 622
 - Software engineering, 222, 223
 - functional decomposition, 232
 - granulate-and-conquer, 232
 - intra-relationships vs. interrelationships, 233
 - structured programming, 233
 - system analysis, 233, 234
 - system design, 234
 - system implementation, 234, 235
 - user requirements, 233
 - Somulti-granular computing, 312
 - Spam, 245, 472–476
 - SPARQL query language, 156
 - Special fuzzy functions models (SFFM), 92, 93
 - Splunk-based SSAS project, 629
 - Spyware, 243
 - Standardization principle, 347
 - Standard rules/practices in numerical computing, 694–695
 - Star-extension, 334
 - Stochastic optimization, 779
 - Stochastic systems, 564–565
 - continuous stochastic systems, 565–578
 - discrete stochastic systems, 578–581
 - reducibility, 581–584
 - uncertainty systems with one dimensional random variable and representations, 584–590
 - unification on uncertainty systems, 590
 - Stone representation theorem, 508
 - Strict-evaluation, 343
 - Strongness, 113
 - Strong supplementation principle (SSP), 505
 - Structural dimensionality, 268
 - Subdeterminants, 11–12, 14
 - Subdistributivity, 701
 - Submatrix, 11–12
 - Sub-unital commutative-ring, 327
 - Sunaga, Teruo, 702
 - Support, 51
 - Support-vector machines (SVM), 731
 - Supremum, 328
 - Suspicious DNS activity, 635
 - Symbolic computation, 891
 - Symmetric triangular fuzzy number, 760
 - Syntax, 529
 - System-based semantics, 5
 - System vulnerabilities, 632
- T**
- Takagi–Sugeno Fuzzy Rulebase (TS-FRB) structure, 80, 92
 - Takagi–Sugeno fuzzy systems, 728–732
 - Takagi–Sugeno method, 844
 - Takagi–Sugeno system, 851
 - Takagi–Sugeno-type model, 850, 851
 - Tangent, 338
 - Taxonomy, 421
 - Taylor expansion, 606, 610
 - Technology-assisted review (TAR), 469
 - Ten-fold cross validation, 398
 - Ternary continuous function, 573, 574
 - Third normal form (3NF), 184
 - Three-dimensional random vector, 572
 - Time-dependent reliability analysis, 58
 - Time-to-live (TTL) value, 635
 - Tolerance relations, 280, 285
 - Topological Boolean algebra, 529, 540
 - Topological fields, 699
 - Topological neighborhood system (TNS), 205
 - Topological quasi-Boolean algebra, 530, 543
 - Topological quasi-field of sets, 544
 - Topological rough algebra, 530, 543–545
 - Topology, 225
 - Total covering rule, 387
 - Totally ordered field, 333
 - Total-ordering, 327
 - Traditional mathematics, 891
 - Traditional statistics, 895
 - Training data, 390
 - Transaction processing, 180
 - Transference, 347, 348, 352, 353
 - Transfer Principle, 330, 334, 347
 - Transitive closure, 555
 - Trapezoidal fuzzy interval, 760
 - TRECVID 2005's test dataset, 320
 - Triangle fuzzy sets, 612
 - Triangle wave function, 612
 - Triangular fuzzy number, 760
 - Triple, 326, 344
 - Triumph of naive set theory, 342
 - Truth preserving property, 314
 - TSK fuzzy model, 839

Twitter, 240
 Two variable function, 335, 342
 Type-1 fuzzy logic systems, 744
 Type-1 fuzzy sets, 743
 Type 1 fuzzy system, 78
 Type-2 and uncertainty
 applications, 753
 data and fuzzy technique, 744
 generalized type-2 fuzzy sets, 747
 imprecision, 744, 745
 intersection of type-2 Fuzzy sets, 747
 interval singleton T2 FLS, 751
 interval type-2 Fuzzy sets, 748, 749
 IT2 FLS, 750, 751
 IT2 FLSS, 748
 representation theorem, 749
 type-1 sets, 746
 type-2 fuzzy sets, 745–747
 type-reduction (TR) methods, 752
 vagueness or linguistic uncertainty, 745
 Type-2 fuzzy logic system, 743, 751
 Type-2 fuzzy model, 834
 Type II fuzzy sets, 165
 Type-2 fuzzy systems, 79, 745
 Type-reduction (TR) methods, 752

U

Ultrafilter, 332
 Uncertainty, 183, 360, 369–371
 theory, 215
 Uncertainty system, 563, 565, 569, 572, 573, 578, 580
 unification on, 590
 with one dimensional random variable and
 representations, 584–590
 Undetermined form, 329
 Unified modeling language (UML), 675–679
 Unital commutative-ring, 327
 Unitality law, 326
 United States law enforcement, 673
 Universal set, 551
 Universal turning machine (UTM), 192
 University of Wisconsin Mathematics Research Center
 (MRC), 703
 Unrestricted fusion (UF), 505
 Upper approximation, 421, 878–881
 Upper membership function (UMF), 748
 Ur-element, 344
 User-based semantics, 5
 User Diagram (UD), 677, 680–683
 Usuality constraint, 37

V

Vagueness/linguistic uncertainty, 745
 Vanishing sequence, 329, 331

Variable-consistency dominance-based rough set approach
 (VCDRSA), 130–131
 Variable precision rough set model (VPRSM), 520
 absolute set approximation regions, 522–523
 approximation regions in, 521–522
 set frequency and conditional frequency, 520–521
 V-imprecisation, 32
 Venn diagrams
 accuracy and coverage, 378
 of defined rules, 380
 of exclusive rules, 379
 of negative rules, 379
 of positive rules, 378
 for probabilistic rules, 378
 Verdegay, 773
 Veristic constraint, 37
 Vladik Kreinovich, 706
 Voice over Internet Protocol (VoIP) communication
 receivers, 736
 von Neumann architecture, 192
 Vulnerability(ies), 632
 data, 626
 scanner, 626
 Vulnerability-related policy violation, 634

W

Warmus, M. (Mieczyslaw), 702
 Wavelet transformation, 493
 Weak supplementation principle (WSP), 505
 Weather prediction, 913
 Web ontology language (OWL), 153–155
 Weighted means, 717
 Weighted ordered weighted averaging (WOWA) operator,
 719–720
 Well-ordering property, 347
 Whitehead, A.N., 507
 Wiener process, 342, 353
 Worm propagation, 242
 Wrapping effect, 701, 704

Y

Young, Rosalind Cecily, 702

Z

Zadeh, L.A., 861
 Zadeh implication, 845
 Zadeh's extension principle, 899–901
 Zero-knowledge proof of knowledge (ZKPK) protocols,
 642
 Ziarko's variable precision model (VPRS), 378
 Zimmermann's approach, 758
 Zimmermann's method, 773